XI Guidelines for Poststop Analysis

In Chapter 1 we presented two research questions related to the analysis of vehicle stop data:

- Does a driver's race/ethnicity have an impact on vehicle stopping behavior by police?
- Does a driver's race/ethnicity have an impact on police behaviors/activities during the stop?

Six chapters in this book have been devoted to the first research question (Chapters 5 to 10), and just this one focuses on the second. This difference in coverage is not an indication of their relative importance, however. In fact, stakeholders have expressed concern and some research has indicated that poststop activities may be at even greater risk of racially biased policing than stop decisions (see, for instance, Harris 2002; Langan et al. 2001) and may have greater negative consequences.

Police can exercise considerable discretion in making poststop decisions (for instance, whether to request consent to search, what disposition to give) and, as we've discussed in other chapters, high-discretion decisions are at greater risk than low-discretion decisions for racial bias. Poststop decisions may result in significant costs for motorists. For instance, searches intrude on motorists' liberty and privacy, they produce fear and even embarrassment, and they "mark" the person as a suspect. A decision to give someone a ticket rather than a warning will have primary costs in the form of fines and potential secondary costs in the form of insurance rate increases. Six chapters pertain to stop analyses because more varied methods have been developed by researchers for examining stops by police than poststop behavior by police. In short, much more attention has been paid to stop analyses than poststop analyses.

This chapter concentrates on the poststop activities most commonly examined by jurisdictions: searches and stop dispositions (the officer's decision to arrest, ticket, warn, or provide no disposition). The principles we explain for analyzing these activities provide guidance for the analysis of other aspects of the stop (for example, length of stop and whether a person was asked to exit the vehicle).

GATHERING DATA TO USE FOR POSTSTOP ANALYSES

For analyzing poststop data, the researcher needs to consider which subsets of data to use. In order to explain this point, we begin by reviewing analysis of stop data. To analyze "who is stopped," a researcher can legitimately analyze a subset that includes only (1) proactive stops and (2) stops in which the officer knew the race/ethnicity of the driver at the time the stop was made (see Chapter 4). Two subsets of data (reactive stops and stops in which the officer could not discern the demographic characteristics of the driver) can be removed because, if racial/ethnic bias influences stop decisions, it will do so in the situations where the officer exercises his or her discretion regarding whom to stop (proactive stops) and when the officer can discern (or thinks he or she can discern) the race/ethnicity of the driver.[1] Once the vehicle is stopped, however, these distinctions are moot since the officer is again proactive (making his or her own decisions) and can perceive the driver's race/ethnicity during

[1] See Chapter 4, pages 61–70, on the rationale for using these subsets of data.

the face-to-face interaction with the driver.[2] As a result, the researcher conducting poststop analysis can include data from reactive stops as well as proactive stops, and data from stops in which the officer could not discern the race/ethnicity at the time the decision to make the stop was made, as well as stops in which race/ethnicity could be discerned.

To remove variation in the level of stopping activity by police across geographic areas of a jurisdiction, subarea analyses were proposed for stop data. The analyses of poststop data should be conducted within these same selected subareas. As noted in Chapter 2 (see Table 2.2), more intense stopping activity in a high minority area might produce results for the jurisdiction indicating bias even when none exists. Similarly, misleading results could be obtained in the analyses of persons police search if subarea analyses of search data are not conducted. The volume of searching behavior can vary across geographic areas, and those areas also can vary by demographic makeup. Therefore, subarea analyses are needed to get an accurate picture of whether police bias influences searching activity in a jurisdiction. The dispositions received by drivers also can vary by geographic location of the stop. More serious dispositions for traffic violations, for instance, may be selected by police in areas with high rates of accidents or numerous complaints by residents about speeding. Conducting subarea analyses will control for some of the area-specific factors that legitimately can influence an officer's decision to arrest, ticket, or warn a motorist or provide no disposition.

ANALYZING SEARCHES

As noted earlier, this chapter focuses on the data analysis for two poststop activities by police: searches and choice of disposition. The resources for the former are described below.

---

2 Information regarding race/ethnicity might come from the driver's appearance, specification of race/ethnicity on the driver's license, name on the driver's license, and so forth.

Resources Required

For effective analysis of search data, jurisdictions must make sure that officers collect certain information on the forms they fill out. The form should include an item indicating whether or not a search was conducted. If officers are instructed to report all searches—not just searches of drivers—the form should indicate what was searched (for instance, vehicle, driver, passengers). In addition, the form should solicit information on the legal authorization for the search. Possible responses include probable cause, consent, reasonable suspicion that a person is armed, incident to arrest, warrant, inventory, and probation/parole waiver.[3] Many data collection forms include "plain view" as a type of search. Technically, plain view is not a "search" since it falls outside the constitutional definition. It is, however, a valid basis for a seizure, and therefore it is appropriate to list "plain view" on the forms. Like plain view, the use of a canine to detect drugs or bombs from outside the vehicle is not technically a search but is appropriately included on some forms.

With regard to consent searches, the forms completed by agencies have differed. Some agencies only record the persons subject to a consent search; these agencies simply include consent in the list of authorizations. Other agencies, however, also record all persons whose consent to search was requested; these

---

3 Agencies should carefully train their officers to ensure that they are consistent and accurate in completing the "authority to search" portion of the data collection form. When we discuss hit rates below, we note the importance of being able to identify the group of evidence-based searches (based, for instance, on probable cause or reasonable suspicion). This becomes complicated when, in an encounter, there are multiple bases for a search. As a rule of thumb, officers should be asked to record the primary reason for the search. A common error occurs when the officer conducts a probable cause (or reasonable suspicion) search, finds something, makes an arrest, and then records "search incident to a lawful arrest" as the authority. This does not correctly reflect the basis of the search that produced the contraband/evidence. Additionally, the form should collect data in such a way that the analyst can determine whether a search occurred that led to an arrest or an arrest occurred that led to a search (Farrell et al. 2004, 28).

agencies include a separate item on the form: "Did you request consent to search? Yes or No." We recommend that jurisdictions collect information about from whom consent was requested by officers and about who was subject to a consent search (because they consented). Below we convey how this comprehensive information can be used to assess search decisions.

An item should be included on the form to indicate search results are either "positive" (something found) or "negative" (nothing found). An agency might decide to include information regarding what the officer expected to find and what was recovered (categories might include currency, weapon, stolen property, illegal drugs, and other) and amount recovered (for instance, amount of drugs, number of weapons).

The search data collected on the forms can be assessed in two general ways: researchers can calculate the "percent searched" for each racial/ethnic group, and researchers can calculate "hit rates" (the percent of searches in which the officers find something).

"Percent Searched" Data

"Percent searched" measures are produced by calculating for each racial/ethnic group the percentage of stopped drivers who are searched. If during a specified period, 100 minorities were stopped in their vehicles and 20 of them were searched, then the percent searched is 20 (20/100 x 100). If 200 Caucasians were stopped in their vehicles and 35 of them were searched, then the percent searched is 17.5 (35/200 x 100).

In many reports and frequently in press coverage, these percentages are used erroneously to draw conclusions regarding racial bias. Analysts, stakeholders, reporters, and even expert witnesses commonly report that higher proportions of stopped minorities were searched compared to stopped Caucasians and mistakenly conclude that this indicates bias on the part of police. Such conclusions are not supported by "percent searched" information.

Misuse of "Percent Searched" Data

The calculation provided above measures searches relative to who was stopped. By calculating for each racial/ethnic group the percentage of stopped people who are searched in a particular jurisdiction, researchers can determine disparity in police decisions to search. But "percent searched" data cannot identify the cause of that disparity or, relatedly, whether or not the disparity is justified.

A key component of stop data analysis is to develop a comparison group that represents the people at risk of being stopped by police absent bias. All of the benchmarks discussed in earlier chapters, regardless of benchmark quality, help to identify "disparity." A major gauge of benchmark quality, we argued, was the extent to which the benchmark was able to rule out nonbias causes of that disparity. In search data analysis, researchers want to know why it is that in most jurisdictions police do not search the same proportions of the stopped drivers in all racial/ethnic groups. In fact, in a majority of reports reviewed by the author, African Americans and Hispanics are searched in higher percentages than Caucasians.[4] Why is it, we might ask, that a jurisdiction whose officers search 15 percent of stopped Caucasians, don't similarly search about 15 percent of the stopped African Americans, 15 percent of the stopped Hispanics, and 15 percent of the stopped "other" racial groups? One explanation is bias. Another explanation—an "alternative hypothesis"—is that racial/ethnic groups are not equally represented among the people at legitimate risk of being searched by police absent bias. Drawing conclusions regarding the existence or lack of bias using stops as the benchmark for searches, is based on a faulty assumption: all stopped people are at equal legitimate risk of being searched. This assumption is contrary

---

4 See, for instance, Schafer et al. forthcoming; Cordner, Williams, and Zuniga 2001; Cox et al. 2001; Decker and Rojek 2002; Spitzer 1999; and Zingraff et al. 2000.

to law and policy. The police are not authorized to search every person they stop.[5]

The people at legitimate risk of being searched by police are the ones who give police cause for a search. “Cause” for a search varies by search type. For instance, an arrest is the “cause” for a search incident to a lawful arrest; of course, probable cause is the “cause” for a probable cause search.[6]

Correct Use of “Percent Searched” Data

The information about percent searched can legitimately be used to describe police searches in the jurisdiction; conclusions can be drawn about disparity but not the cause of that disparity. Agencies might use the “percent searched” measure to describe searches not only in terms of racial/ethnic groups, but also for gender groups and/or for gender groups within racial/ethnic groups (Figure 11.1).[7] This figure illustrating data for a hypothetical jurisdiction indicates that 16 percent of the Caucasian males who were stopped by police were searched. Corresponding figures for African American males, Hispanic males, and “Other” males were 24, 21, and 15 percent, respectively.[8] Similar information is provided for the females who were stopped by police.

---

5 Rules for when police can search are set forth in court decisions, in federal and state constitutions, and in agency policy. Most relevant here are the rules governing searches after vehicle stops. Police can search anyone who has just been arrested (“search incident to a lawful arrest”), search a car when it is being impounded (“inventory search”), retrieve evidence/contraband in plain view in an occupied automobile, frisk (conduct a “pat down” of) a detained person whom the officer reasonably believes is armed and dangerous, search a person or automobile based on probable cause that contraband/evidence will be found, and search a person or automobile if the person consents.

6 Consent searches, which we discuss later in this chapter, are comprised of searches based on cause (based on some level of evidence up to and including probable cause) and searches based on no articulable evidence supporting suspicion.

7 Agencies with a low number of searches will not be able to break the search information down into so many categories.

8 Schafer, Carter, and Katz-Bannister (2004) presented their results by categories of searches (for instance, consent searches, incident to arrest, plain view), gender, and race.

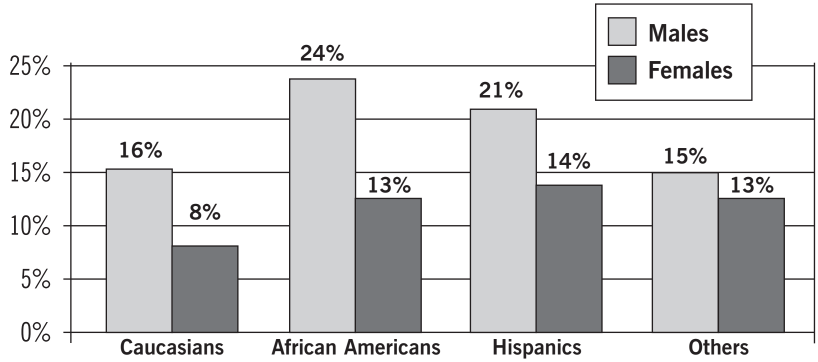

Figure 11.1. Searches as a Percentage of Vehicle Stops, by Race/Ethnicity and Gender of Detained Group, Hypothetical Jurisdiction

Although "percent searched" information such as that provided in Figure 11.1 cannot prove the existence or lack of bias, it is still important and instructive. It can convey disparity, and therefore it is worthy of review and discussion (see Chapter 13). It may even provide sufficient basis for intervention. An agency that produced data similar to the data in the figure might want to consider why a full 24 percent of detained African Americans were searched by its officers. The law enforcement executive might want to examine officers' behavior more closely to understand this finding, or the executive might acknowledge the possibility that bias may be influencing these high-discretion decisions and implement reforms.

Search "Hit Rates" Defined

From search data researchers can calculate "hit rates," in addition to the "percent searched" information described in the previous section. A hit rate is the percent of searches in which the officers find something upon the people being searched. Officers might find contraband (for instance, drugs, illegal weapons) or other evidence of a crime. Hit rates can be calculated for each racial/ethnic group, and sometimes (if the number of searches is high enough) for racial/ethnic groups by gender and even age.

Assume for the sake of example that, during a specified reference period, police in an agency searched 100 of the stopped Caucasians, 80 of the stopped African Americans, and 60 of the stopped Hispanics and found evidence on 10 of the Caucasians, 4 of the African Americans, and 4 of the Hispanics. The hit rates would be 10 percent for Caucasians (10/100 x 100), 5 percent for African Americans (4/80 x 100), and 7 percent for Hispanics (4/60 x 100).

Search Hit Rates as a Problematic Measure of Criminality

What do hit rates convey? This question has been widely debated. Some researchers and commentators claim that hit rates provide a measure of criminality (see, for instance, Harris 2002; Alschuler 2002). Those making this claim argue that a finding of equal hit rates in searches of Caucasians and minorities is evidence that minorities are not more criminal than Caucasians, despite widespread perceptions to the contrary. Claims such as this one are usually made in the context of arguing that racial profiling is not an effective law enforcement tool.

Other researchers dispute the claim that hit rates convey information about criminality (see Banks 2003; Harcourt 2003; Knowles, Persico, and Todd 2001). The problem of using hit rates to measure criminality is analogous to the problem of using arrest rates to measure criminality. Hit rates and arrest rates are affected by citizen behavior and police behavior. As discussed in Chapter 10, the demographic profile of people arrested in a jurisdiction reflects two factors: (1) persons who commit crime (citizen behavior) and (2) persons whom the police identify and target for arrest (police behavior). Similarly, search hit rates reflect not only the people within each racial group who are carrying evidence/contraband; they also reflect police choices regarding whom to search. Hit rates could serve as a legitimate measure of criminality under only one circumstance: if the police or a researcher randomly selected people for search (Banks 2003). If we randomly selected people in a jurisdiction for search and found that 8 percent of the Caucasians and 8 percent of the racial/ethnic minorities were carrying contraband, we could argue that the carrying of contraband is equal across racial/ethnic groups. Of course, neither researchers nor police can conduct this hypothetical study for various legal, moral, and practical reasons.

Search Hit Rates as a Measure of Unjustifiable Disparity

Search "hit rate" data cannot provide sound information on the criminality of the general population, but—like "percent searched" data—it can indicate disparity or lack thereof in police decisions to search. Moreover, for some types of searches, hit rates can indicate, not just disparity, but "unjustifiable disparity" in police decisions to search. For this subset of searches, search hit rates can rule out (not definitively but with an acceptable degree of confidence) the alternative hypotheses (hypotheses that factors other than bias influence police behavior). An economic theory called the "outcome test" will help us understand how, for some searches, the hit rate analyses can measure unjustifiable disparity. We begin by explaining the theory behind the outcome test.

The Theoretical Basis of the "Outcome Test"

The outcome test can show whether decision makers used different criteria for different groups. This test can be applied, however, only when decision makers claim that their decisions are based on the probability of a particular outcome. The outcome test, first proposed by Nobel Prize-winning economist Gary S. Becker (1993), was applied by him to outcomes related to money lending. It also can be applied to certain types of searches.[9] The outcome test focuses on the pool of people that the decision maker deems qualified for a loan or a search.

---

9 Knowles, Persico, and Todd (2001) were the first to apply the outcome test to data on vehicle stops.

Yale University Professor Ian Ayres (2001, 404) reports that "outcome tests can provide powerful evidence of when a particular kind of decision making has an unjustified disparate impact" on a particular group, such as a racial group. Ayres reviews Becker's application of the outcome test to money lending decisions.

Assume a bank claims to make decisions regarding who will get loans based only on criteria that pertain directly to the likelihood that the borrower will be able to pay the loan back. To make these decisions, the bank analyzes the borrower in terms of assets, credit history, and other factors directly related to the bank's goal: loan repayment. If the bank applies this criteria (probability of loan repayment) equitably across all racial/ethnic groups and does not consider race/ethnicity as a factor in itself, then the default rates should be equal across groups.[10] In other words, if racial/ethnic groups are evaluated along the same criteria—those criteria related to likelihood of loan repayment—then they should succeed in their loan repayment at the same rates. If, in fact, the minority borrowers default on their loans at a lower rate than their Caucasian counterparts, researchers can infer that minority borrowers were held to a higher standard by those deciding to make the loans. Or stated another way, from this lower rate of default for minorities, researchers can infer that "qualified" minority borrowers were denied loans.

The above example pertains to the potential differential allocation of benefits (for instance, loans) across racial groups. As Ayres reports, the same test can be used to assess a decision maker's allocation of detriments (that is, distribution of "bad things"). Ayres (2001, 405) writes, "If we find that in distributing a detriment that the decision maker effectively accepts poorer outcomes from minorities than from whites, we may infer there to be a class of minorities that might have avoided the

---

10 We acknowledge that this example simplifies loan decisions. In the real world, loan decisions are not simply whether to provide or deny a loan, but involve setting interest rate levels. Bias might be manifested, not just in decisions to give or deny loans, but more subtly in the form of higher interest rates.

detriment." An example of a "detriment" is a police search. He explains, "if we find police search decisions are systematically less productive with regard to minorities than with regard to whites, we might infer that search decisions have an unjustified disparate impact in subjecting undeserving minorities to being searched" (Ayres 2001, 406).

Such a finding—that there is a lower rate of search hits for minorities than for Caucasians—indicates that different standards were utilized in selecting Caucasians and minorities for searches. Specifically, the implication is that a lower standard of proof was applied to searches of minorities than to searches of Caucasians.

The outcome test does not focus on whether different proportions of minority and Caucasian applicants make it into a pool of loan recipients. More Caucasians than minorities might meet legitimate, unbiased qualifying criteria for a loan. With regard to searches, different proportions of Caucasians and minorities might meet even the legitimate, unbiased criteria for a search. As noted earlier, the outcome test focuses on the pool of people that the decision maker "deemed qualified" for a loan or a search.

Another way to restate the example is by using a hypothetical construct, "units of evidence." Imagine an officer who searches all minorities he detains for whom he has 50 units of evidence that they are carrying contraband or other evidence. He searches all Caucasians he detains for whom he has a corresponding 80 units of evidence. He has set a lower standard for searching minorities compared to Caucasians. The result will be that he is "wrong" more often with his minority searches; the officer is less likely to find evidence on the minorities, because he settled for a low level of evidence to initiate the search. He will have more "hits" in his searches of Caucasians because he didn't search them unless he was highly confident that they were carrying contraband/evidence.[11] This produces a lower hit

---

11 A "wrong" decision does not imply that the search was unjustified; similarly, a "hit" does not imply that the basis for the search was legitimate.

rate for minority searches. As Ayres explains (2002, 133), "A finding that minority searches are systematically less productive than white searches is accordingly evidence that police require less [evidence] when searching minorities."

A Unique Virtue of the Outcome Test

Most social science methods are susceptible to "omitted variable bias." The researcher cannot identify and/or measure all the variables that might affect the outcome being studied. The outcome test, however, is not susceptible to the traditional omitted variable bias concern (Ayres 2001, 2002). In the outcome test, we do not need to know all the factors that the banker considered nor all the factors that the police officer considered in order to isolate cause (Ayres 2002; Knowles, Persico and Todd 2001). We merely need to know that they claim their decisions are made only on the probability of some outcome and not on other variables.

The outcome test is not impeded by the possibility that a variable omitted by the researcher influences the outcome. In fact, as Ayres (2002, 133) explains, "the outcome test intentionally harnesses omitted variable bias to test whether any excluded (unjustified) determinant of decisionmaking is sufficiently correlated with the included racial characteristics to produce evidence of a statistically significant disparity. Any finding that the police searches of individuals with a particular characteristic (such as minority status) induce a systematically lower probability of uncovering illegality suggests that police search criteria unjustifiably subject that class of individuals to the disability of being searched." Such results indicate, in Ayre's terminology, an "unjustified disparate impact" of search decisions.[12]

---

12 There is debate among analysts and scholars regarding whether the researcher should look at the combined outcomes of "percent searched" and "hit rates" (based on the outcome test) to draw conclusions about racial bias. We argue that percent searched data do not provide information about bias either alone or when viewed in conjunction with hit rate results.

Unjustified Disparate Impact or Disparate Treatment?

What exactly does "unjustified disparate impact" mean? As noted earlier, the outcome test can show that decision makers used different criteria for different groups. "Unjustified disparate impact" means that the outcome for one group is different than the outcome for another group, no justification exists for this disparity, but no specific cause or motive can be proved. In other words, "unjustified disparate impact" is disparity that cannot be explained by legitimate causes. It is a red flag for racial bias, but it does not prove racial bias.

Ayres makes a careful distinction between "unjustified disparate impact" and "disparate treatment" or bias. "Disparate treatment" implies that the decision maker's decision was inappropriately based on race/ethnicity; it sets forth a specific cause for the unjustified disparate impact. In the context of searches, to find "disparate treatment" implies that any identified disparity is the result of decisions inappropriately influenced by race—for instance, a reduced standard of proof for minority searches. In contrast, "unjustified disparate impact" is neutral as to motives or cause—because they cannot be discerned through the evidence.

Lower hit rates for minorities are cause for concern. These results are a warning signal requiring the serious attention of law enforcement agencies. There are some explanations, however, other than bias for these results. They will be discussed later. First, we describe the types of searches to which the outcome test applies.

The Outcome Test and Evidence-Based Searches

For all types of searches, hit rates provide descriptive information regarding whether or not there is disparity in productivity. If, for instance, 22 percent of the searches incident to arrest of African Americans produced hits compared to 30 percent of the searches incident to arrests of Caucasians, we know that the Caucasian searches of this type are more productive. This is valuable information for further exploration even though we cannot determine whether or not bias is the cause of this disparity.

For certain searches—the ones that meet the assumptions of the outcome test—researchers can gain additional information: they can say with reasonable confidence that any identified disparity is unjustified and likely (but not certainly) caused by bias. The outcome test applies in narrow circumstances: when decision makers claim that their decisions are based only on the probability of a particular outcome. As Ayres (2002, 134) explains, "The decisionmaker in an outcome test by her own decisions defines what she thinks the qualified pool is, and the outcome test then directly assesses whether the minorities and nonminorities so chosen are in fact equally qualified." The bankers will claim that they make loan decisions based only on the probability of default. The corresponding circumstance for police is when they make searches based on the probability of finding contraband/evidence.[13] This is true when the police conduct probable cause searches, frisks for weapons, searches based on "plain view" or drug odors, and, arguably, canine alert searches.[14] We will refer to these types of searches as "evidence-based searches." The requirement of the outcome test (decisions must be based on the probability of a certain outcome) is not met with other types of searches, such as searches incident to a lawful arrest, inventory searches, or warrant searches.[15]

---

13 Note that hit rate analysis has been applied to other criminal justice decisions. This includes bail decisions in which the judge makes a decision regarding release based on the likelihood that the defendant will appear at trial (Ayres 2001) and MDT queries where the officer runs a query based, presumably, on a belief that it will turn up negative information about the driver or car. See our discussion in Chapter 10 of the research of Meehan and Ponder (2002).

14 Whether searches based on probation/parole waiver are evidenced-based searches may vary by agency and/or by individuals within agencies. If any officers within a department take advantage of this legal authority to search regardless of their assessment of the probability of finding evidence, then this category of searches does not meet the assumption of the outcome test.

15 Later we will explain why hit rate analysis based on the outcome test cannot usually be conducted on consent searches.

Table 11.1. Evidence-Based Search "Hit Rates," by Race/Ethnicity, Gender, and Age, Hypothetical Jurisdiction

Race/Ethnicity	Female		Male	
	≤24	25+	≤24	25+
Caucasians	17%	14%	15%	16%
African American	13%	15%	8%	15%
Hispanic	15%	16%	6%	17%
Other	15%	13%	14%	15%

Source: Based on a table in Council on Crime and Justice and Institute on Race and Poverty (2003, 29).

Researchers can calculate hit rates for all types of searches (grouped together or separated by type) and describe whether disparity exists. A separate analysis of the searches that meet the assumptions of the outcome test—evidence-based searches—could then be conducted that will allow the researcher to state with a reasonable degree of confidence that bias is or is not indicated. Refined assessments would compare hit rates for racial/ethnic groups within specific subcategories of the evidence-based searches, such as categories based on the justification for the search (for instance, probable cause, plain view) or, if the relevant information is available on the form, on the basis of what the officer reports he was seeking (for instance, drugs, weapons).

Table 11.1 provides sample results showing hit rates for evidence-based searches for groups defined by their race, age, and gender. These hypothetical data indicate that hit rates for evidence-based searches of young minority males are lower than for any other group.[16] Results such as these should prompt law enforcement agencies to examine their searches more closely and/or implement interventions to reduce this apparent bias in searches.

---

16 Again, a small number of searches may preclude breakdowns of the data within categories such as race, gender, and type.

Nonbias Explanations for Differential Hit Rates Based on the Outcome Test

Why can agencies legitimately consider lower hit rates for minorities for evidence-based searches a red flag for potential racial bias? They can come to this conclusion because the outcome test reduces significantly the nonbias explanations for the outcome. Therefore, these findings are sufficient grounds, at the very least, for further exploration by a department. But, in fact, nonbias-related factors can produce lower hit rates for minorities. For this reason, Ayres reports that hit rates based on the outcome test cannot prove "disparate treatment." They only can indicate "unjustified disparate impact."

Various circumstances can produce lower hit rates for minorities even when police decisions are void of bias. Ayres refers to this as a false positive, because the hit rate results indicate a problem when, in fact, the problem does not exist. Ayres refers to the circumstances that might produce a false positive result as the "subgroup validity problem."

False Positive Results and the Subgroup Validity Problem

In relation to searches, the subgroup validity problem occurs when there is a particular characteristic linked to the probability of carrying contraband/evidence that is valid for one subgroup but not for another. If the decision maker uses this characteristic in decision making, racially disparate outcomes may result that do not reflect bias.

Ayres explains the subgroup validity problem with the following example. The wearing of a particular type of hat is strong evidence of drug possession when worn by Caucasians, but it is not evidence of possession when worn by minorities. "In the extreme," states Ayres (2002, 139), "imagine that 100 percent of whites wearing this cap possess drugs, and 0 percent of minorities wearing this cap possess drugs." If the police use hat wearing as part of the "totality of the circumstance" in justifying searches, the result will be a lower hit rate for the minorities. The lower hit rate for minorities will result because the hat clue will be strong only for Caucasians and not for minorities. The police will, in effect, have better evidence for detecting Caucasian drug carrying than minority drug carrying because they have a convenient visual cue for one group but not the other.

A more realistic example pertains to nervousness as a clue for illegal carrying. Police sometimes cite nervousness as one factor in the "totality of the circumstances" to justify a detention and/or search. What if nervousness is linked to carrying contraband/evidence in one racial/ethnic group and not the others? This might be the case, as one practitioner suggested to the author, with regard to Hispanics. In that officer's experience, Hispanics were more likely to be nervous around police regardless of whether they were carrying contraband/evidence. He explained that many of the immigrant Hispanics in his southwest U.S. jurisdiction had had bad experiences with police in their native countries. Some of them were illegal aliens, fearing deportation. This led them to be fearful in the presence of police. According to this example, nervousness is a good indicator of carrying contraband/evidence for Caucasians and maybe for African Americans but not for Hispanics. If police apply the nervousness criteria to all demographic groups, they will produce lower hit rates among Hispanics because the nervousness criteria will not be as effective in predicting carrying of contraband/evidence for this group.

The subgroup validity problem may be counteracted by astute and effective policing. Arguably, astute police would come to recognize that blue hats are associated with Caucasians carrying drugs and not with minorities carrying drugs. Police would adjust their search decisions accordingly. Astute police aspiring to effective searches would come to recognize that nervousness is not as viable a clue for carrying for Hispanics as for non-Hispanics, and they would tailor the use of nervousness to justify search decisions accordingly.

False Negative Results: Racial Bias Undetected by Hit Rate Analysis

Above we described the circumstances in which the outcome test results may indicate bias (that is, show lower hit rates for minorities) when a nonbias reason exists. Ayres (2002) notes that there also are circumstances in which the outcome test will not detect existing racial bias. This "false negative" will occur, he reports, when police use race as a proxy for criminal activity and their predictions prove correct (see also Borooah 2001). Ayres (2002, 135) describes these police decisions to search as "based on valid statistical inference." (Importantly, he does not use "valid" to imply these inferences are legitimate. He believes they are not. He uses the term to refer to statistical inferences that turn out to be correct.) Ayres (2002, 135) writes: "For example, if police were correct in inferring among some group of otherwise observationally equivalent suspects that minority suspects had a higher likelihood than whites of possessing contraband and used the race expressly as a part of their criteria for searching, then in equilibrium we might not observe lower search success rate for minorities than for whites." Specifically, in the scenario he describes (police "correctly" predict a minority-crime link), the hit rate for minorities could be higher than for Caucasians. Ayres calls this a "false negative" on the premise that making search decisions based on the minority-crime link is a form of "racial profiling."[17] [18]

---

17 This premise, however, is disputable. In *U.S. v. Martinez-Fuerte*, 428 U.S. 543 (1976), the U.S. Supreme Court upheld police use of race/ethnicity in some circumstances as a proxy for criminal activity. In those specific circumstances, the "statistical inference" would not signal racial bias, as least as defined by the Court. Law enforcement agencies are still awaiting clear guidance by the U.S. Supreme Court on this very important and very controversial question.

18 In addition to the subgroup validity problem, Ayres identifies another weakness of the outcome test. He calls it the "infra-marginality problem." This weakness is responsible in part for the test's ability to detect only "unjustified disparate impact" and not "disparate treatment." (continued)

Other Factors to Consider When Interpreting Hit Rates

Above we explained that the subgroup validity problem could produce unequal hit rates when no bias exists and that "statistical inference" could mask biased search decisions. There are several other factors or circumstances that could impact on hit rates; all of these save one reflect other ways that bias might manifest in search-related activities.

Data Quality

Rudovsky (2001), Gross and Barnes (2002), and Fagan (2002) raise the issue of the validity of search data. The issue of data quality was addressed in Chapter 4, and some methods set forth to help jurisdictions detect and remedy intentional and unintentional errors in the stop data gathered by police. While no agency will be able to ensure that its data are 100 percent correct, a strong audit system coupled with supervisors who hold officers accountable for the data they are supposed to submit will improve data quality. Systematic intentional or unintentional errors could have an impact on an agency's search hit rates. For example, if officers consistently fill out search forms for every search of a Caucasian but, in defiance of policy, "neglect" to fill out search forms for many minorities except when contraband is found (Gross and Barnes 2002), the hit rates for minorities will be artificially high. The incomplete data might

---

[18] (continued from previous page) Ayres notes that in the strongest application of the outcome test, the outcome is a continuous variable rather than a dichotomous one. Loan defaults and search hits are both dichotomous outcomes—they either happen or they do not. An outcome that manifests as a continuous variable (that is, "degrees" of outcome achievement) would allow the researcher to focus narrowly on the people most at risk of biased treatment—those "at the margins." Generally speaking, the outcome test is strongest when the researcher can examine activity "at the margins"; with dichotomous outcomes, the researcher can only examine "average outcomes." Ayres makes a persuasive argument, however, that "average" search outcomes for racial groups are "probative of marginal (or threshold) outcome differences" (2002, 138). For more information on the infra-marginality problem, see Ayres (2002, 135-138).

mask the true situation: low hit rates for minorities that could indicate racial bias.

Search Intensity

The nature of the search itself may affect the likelihood of a hit. The more intense a search, the more likely something will be found. Even if officers select people to search without bias, but then search minorities with more intensity, the hit rates for minorities could be inflated. This circumstance, if combined with decisions to search that are racially biased, could produce "false negative" hit rate results. If the officers are using lower standards of evidence to search minorities and searching those minorities more intensely than Caucasians, the lower hit rates that should have resulted from the lower level of evidence could be offset by the intensity of the searches that produced more hits.

"Subsearch Processes"

Gross and Barnes (2002) raise the issue of "subsearch" processes (see also Harcourt 2003; Fagan 2002). Gross and Barnes describe the various activities in which police might engage prior to conducting a search that could help them make more informed search decisions. An officer might, prior to deciding whether or not to search, order driver and passengers out of the car, engage in some level of questioning (for instance, asking driver and passengers separately where they are going, where they are coming from, and so forth), look into the vehicle from outside, bring narcotics-detection dogs to the scene, or even take the driver's pulse (Gross and Barnes 2002). Such activities could help an officer determine whether a search is justified by the totality of the circumstances. If officers engage in these subsearch processes differentially based on the race of the driver, hit rates could be influenced. Imagine some biased officers who have no compunction detaining minorities for lengthy periods while they question them, look into their car windows, and bring search dogs to the scene. This extensive information gathering will help the officers determine whether or not a search is supported by evidence. This could result in more and "better researched" searches of minorities than of Caucasians and could result in higher hit rates for minorities.

Differential Standards of Proof for Searches for Different Crimes
Hit rates based on the assumptions of the outcome test provide information on officers' decisions to search. All of the factors or circumstances described above represent other ways that bias might enter into search-related activities and have an impact on hit rates. Bias might influence form completion, search intensity, or subsearch activities. A final circumstance that could affect hit rates could reflect either biased or bias-free decisions. An agency might set levels of proof to justify searches that vary by the type of crime being investigated. If the perpetrators of these different crimes vary by race, hit rates could be affected.

An example will help to convey the point. Agency Q sets a lower (but constitutional) standard of proof for searches related to suspected weapon possession than for suspected drug crimes. This might occur, for instance, if the agency executive sends a message encouraging a crackdown on weapons crimes, and training facilitates the identification and articulation of evidence that amounts to probable cause. This could produce weapons searches based on "just enough" evidence to produce probable cause. If there is no such message and training pertaining to drugs, drug searches may be based on a higher level of proof. If, in the jurisdiction, Blacks are more likely to engage in weapons crimes than Whites, and Whites are more likely to engage in drug dealing than Blacks, the lower level of proof for the crime committed by Blacks will produce a lower hit rate for Blacks. If the decision to lower the level of proof for weapons offenses was based on the fact that Blacks are the predominant perpetrators, then this circumstance adds to our list of how bias could be manifest in search-related activities and have an impact on hit rates. If, on the other hand, the lower level of proof was a race-neutral decision, no bias is at work.

Implications for Interpreting Hit Rates

As we have seen, bias can influence not only an officer's decision whether to conduct a search but the officer's search-related activities; these biased activities can also have an impact on hit rates. In the final circumstance described in the previous section—where levels of proof are related to crime types and crime types are, in turn, linked to race—the decisions may or may not reflect bias. The interpretation of search hit rates is complicated because there are several nonbias explanations for differences in the rates for racial/ethnic groups. Despite these complications, the challenge of interpreting hit rates is still worth the effort by law enforcement agencies.

Because bias can affect hit rates in various ways, any results indicating outcome-test-based hit rates for minorities and for Caucasians are substantively different should lead to agency action. That is, the results that should lead agencies to take additional steps would be *different* hit rates across racial/ethnic groups, not just lower hit rates for minorities. The chance that this difference will reflect *unjustified* disparate impact is not 100 percent, but it is high. Racial bias may influence police decisions concerning the level of proof required for a search, the amount of data they record on the data collection form, the intensity of the search, or the extent of information-gathering activities before the search.

Those "additional steps" by law enforcement agencies could include expanded collection of quantitative or qualitative data on searches or interventions to eliminate or decrease bias in search decisions. Consider this example of further quantitative assessment. An agency that believes differential hit rates across racial/ethnic groups may be due to differential levels of proof for searches of different crimes could choose to collect information on the various types of crimes that officers suspect when they search and then compare Caucasian and minority hit rates within the distinct crime categories. For instance, the agency might look at hit rates for evidence-based searches broken down by the type of crime suspected (such as violent, weapons, property, drug, and other). Alternatively or additionally, an agency might collect information on the various types of subsearch activities conducted by officers to assess their frequency and whether or not there is disparity in their application.

Instead of, or in addition to, expanded data collection, an agency might decide to hold meetings with community residents to discuss the circumstances listed in this chapter that might be manifesting in the jurisdiction to produce the unequal hit rates. That is, practitioners and residents might discuss the various interpretations of the agency's hit rates in light of information in this chapter and decide on further steps—which could include more data collection or intervention. We say more about these police-resident discussions in Chapter 13.

It may be reasonable for an agency to decide to move right to the intervention stage, even if the data it has collected have not "proved" racial bias. This is particularly viable with regard to searches because of the high degree of discretion on the part of officers associated with several types of searches. As conveyed elsewhere, high-discretion decisions are at greatest risk of racial bias. We discuss some possible interventions in Chapter 13.

The Special Case of Consent Searches

Consent searches are highly discretionary actions, and the more discretion associated with an activity, the more likely it is that bias could be manifested. Because of this fact, consent searches should receive special attention by agencies. They should receive special attention even though researchers are limited in their ability to draw conclusions regarding bias from consent search data.[19]

At first glance, consent searches might be considered evidenced-based searches suited to the outcome test. Presumably, officers initiate these activities based on their belief that they

---

19 We discuss in Chapter 13 some actions agencies can take to reduce the potential for bias in consent searches and other high-discretion activities.

will find evidence.[20] There are important differences, however, between consent searches and nonconsent evidence-based searches. With the nonconsent evidence-based group, the decision of the officer that is evaluated in the outcome test is the decision to conduct the search; in every instance the researcher will know whether or not the officer was right or wrong about whether the person was carrying contraband or other evidence. If the officer endeavors to conduct 100 nonconsent evidence-based searches, he will conduct 100 of them, and the researcher will know for each one whether or not there was a "hit." With consent searches, however, the decision of the officer that is evaluated is the decision to *request* consent to search. The researcher wants to know if the officer, because of bias, requests consent to search from minorities more than from Caucasians. The officer may want to conduct 100 consent searches but be able to conduct only 85 because consent is withheld by 15 people. To evaluate the officer's decision using the outcome test, the researcher would need to know for all 100 people who was and was not carrying contraband or other evidence. This information is known only for 85 of the 100. The researcher cannot assume that the 85 are representative of the 100. It is plausible that the 15 who refused to provide consent are carrying evidence/contraband at a higher rate than the 85 who consented, and it is possible that the relationship between refusal and carrying differs across demographic groups.

Table 11.2 clarifies why the outcome test cannot, in general, be applied to consent searches. Officers in a hypothetical agency asked 50 Caucasians and 50 minorities for consent to search. For purposes of the example, assume that 40 percent of both groups are, in fact, carrying contraband or other evidence, so the hit rates

---

20 That said, it is important to note that officers generally do not need to meet legal standards of "evidence" in order to initiate a consent search. A few agencies do have policies that require officers to articulate at least minimal evidence of criminal activity prior to requesting consent to search even if it does not amount to either reasonable suspicion or probable cause.

are equal. This indicates that the officers are applying the same criteria to Caucasians and to minorities when deciding whether to request consent to search. Indeed, these hit rates are needed if the researcher is to evaluate the officers' decision based on the outcome test. Unfortunately, in the real world, those hit rates are not available: the carrying rate of all the people from whom consent to search was requested is not known. The hit rates are known only for those people who granted the officers the requested consent. For both the Caucasians and the minorities in our example, 40 of the 50 people who were asked for consent granted it. Among the Caucasians, the 10 who denied consent *were not* carrying evidence/contraband, and the 10 among the minorities who denied consent *were* all carrying. The resulting hit rates are 50 percent and 25 percent for Caucasians and minorities, respectively. Researchers who do not consider the potential impact on their results of the "missing (hit rate) data" for the people who refused consent might claim the officers in question are manifesting bias in their decisions regarding requests to consent when, in fact, they are not.

Table 11.2. Consent Search Data Showing Why the Outcome Test Cannot Generally Be Applied, Hypothetical Jurisdiction

Race	Consent Requested			Consent Granted		
	Asked	Carrying	Hit Rate	Granted	Carrying	Hit Rate
Caucasians	50	20	40%	40	20	50%
Minorities	50	20	40%	40	10	25%

In Table 11.2 the difference between the two groups in the carrying-status of the people who denied consent was extreme; all 10 of the Caucasians who denied consent were not carrying, and all 10 of the minorities who denied consent were carrying. In this example the potential differences between groups is probably exaggerated, but researchers cannot assume (without some empirical basis for doing so) that similar proportions of carrying and noncarrying minorities and carrying and noncarrying Caucasians deny consent.

In the above example, 10 of the 50 minorities and 10 of the 50 Caucasians denied consent, producing 20 percent "missing data" (for each group and overall) for the key variable required for the outcome test—the carrying rate (hit rate). The percentage of missing data is too high to examine hit rates and draw conclusions regarding unjustified disparate impact. If the level of "missing data" were lower, a case might be made that the researcher could examine the hit rates for unjustified disparate impact. For instance, if the rate of acquiescence were 99 percent for both groups, a researcher could argue that she or he had sufficient proportions of "hit rate" data to conduct the analysis. Although there is no clear rule of thumb for when the level of missing "consent search" data is sufficiently low to determine unjustified disparate impact, we maintain that a researcher who has at least 95 percent agreement to the consent searches within each racial/ethnic group can analyze the data to identify unjustified disparate impact.[21]

An agency that is not able to conduct an outcome-test-based hit rate analysis on their consent search data can still calculate hit rates. These calculations will not produce a measure of unjustified disparate impact, but they can identify disparity in the relative productivity of searches of Caucasians and minorities. As discussed further in Chapter 13, large disparities, while not proof of biased policing, are worthy of review, discussion, and possibly intervention.

---

21 In order for the researchers to be able to conduct these types of analyses, the law enforcement agency must include on the data collection form an item regarding whether or not the person was asked for consent to search.

The Ongoing Debate on Hit Rates

There is an ongoing debate among researchers and practitioners about hit rates, just as there is continued debate about other aspects of vehicle stop analysis. As mentioned above, some of this debate centers on whether hit rates tell researchers something about criminality or about police decisions. Fagan (2002) and other researchers question whether the outcome test solves the omitted variable problem as Becker and Ayres claim it does. As dialogue continues about hit rates, we expect another area of discussion will center on whether the nonbias explanations for differential outcome-test-based hit rates are sufficiently narrow to justify our claim that low hit rates (for evidence-based searches) are a red flag for bias. This topic of hit rates has piqued the intellectual interest of scholars from diverse fields: economics (Knowles, Persico, and Todd 2001; Hernandez-Murillo and Knowles 2004; Persico 2002; Borooah 2001; Borooah 2002; Chakravarty 2002),[22] law (Banks 2003; Harcourt 2003; Harris 2002; Rudovsky 2001; Alschuler 2002) and criminal justice (Fagan 2002; Engel and Calnon 2004).[23] We expect continued, high-level discussion on this topic.

Other Ways to Examine Searches

The "internal benchmarking" method described in Chapter 8 to analyze stopping behavior by police can be applied to searches as well. For stop analysis, agencies compare stops by individual officers to stops by other similarly situated officers, or they compare stops by a group of officers to stops by other similarly situated groups of officers.[24]

---

22 The economists do not just analyze hit rates to assess police bias. They also examine the costs and benefits (including increased or decreased crime) of using race as a predictor of criminality (see Harcourt 2003).

23 Harcourt (2003) provides a comprehensive, analytical overview of the various perspectives.

24 For instance, they compare officers who are assigned to the same geographic area, the same shift, and who have the same mission (such as patrol). These similarly situated officers are exposed to the same group of people at risk of being stopped by police.

Agencies also can compare similarly situated officers with regard to the percent of drivers searched who are minorities. Figure 11.2 illustrates internal benchmarking with search data. In this hypothetical jurisdiction, between 20 and 30 percent of the drivers searched by officers (Officers 1 through 9 and Officers 11 and 12 in the figure) are minorities. In contrast, 45 percent of the drivers searched by Officer No. 10 are minorities. Officer No. 10 is an "outlier" (in social science terminology), and this officer's decisions to search should be reviewed by the department to see if bias is influencing them.

Internal benchmarking could be conducted within search types (for instance, separate analyses for consent searches and warrant searches), within groups of searches (for instance, within high- and low-discretion categories), or for hit rates. The analyst applying this method to searches would follow the recommendations in Chapter 8 with regard to matching officers or units, conducting the analysis, drawing conclusions from the results, and taking appropriate action.

In their report for the San Antonio Police Department, the Lamberth consulting team assessed disparities in search data and then conducted what we refer to as a "qualitative analysis of quantitative data." After summarizing the challenges associated with benchmarking search data, Lamberth (2003b, 43) explains:

> The proportion of stops of minorities typically varies by area of the city, as does the proportion of searches of minorities. Some areas of the City have heavier deployments of police than do others based on such factors as crime, citizen calls for service and the like. Some types of deployments, particularly those aimed at reducing crimes plaguing a specific area may have guidelines to seek to search more aggressively than do regular patrol deployments. Thus, it is not a simple matter to decide upon an appropriate benchmark nor is it an easy task to quantify that benchmark.[25]

---

[25] This report is available on the PERF website at www.policeforum.org.

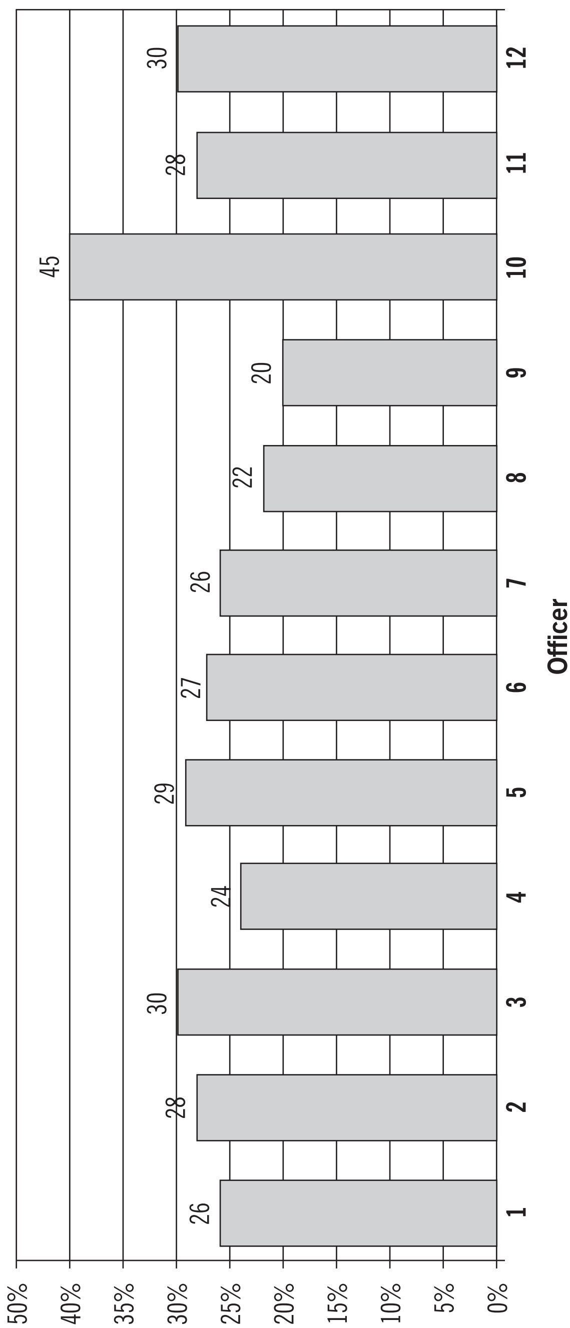

Figure 11.2. A Comparison of Twelve Officers Who Are Similarly Situated ("Matched"): Percent of Drivers Searched Who Are Minorities

The search data are the "numerator" data, a term explained in Chapter 4. The challenges that Lamberth describes are related to matching this data to an appropriate benchmark (the "denominator" data). In fact, a match cannot be made. Lamberth (2003b, 43) writes, "Instead of attempting to specify a benchmark, as we have with the stop data, we will discuss searches in the context of some of the variables that affect them."

In their qualitative analysis, the Lamberth team focuses on consent searches because, as high-discretion activities, they are vulnerable to bias and because the team found that minorities were considerably overrepresented among drivers subject to consent searches relative to their representation among those stopped. The team discusses the search results in the context of specific subareas of the city that vary with regard to the nature and extent of crime, level and type of deployment, and the proportion of residents on probation or parole. Definitive conclusions about racial bias cannot be drawn from the data (all of the alternative hypotheses are not addressed), and the team is careful not to draw definitive conclusions. The team's findings, however, are constructive because they shed light on search activity by police. A qualitative analysis of quantitative data could be conducted on search information by the researchers or be one component of the discussions conducted by police and citizens, as we recommend in Chapter 13.

Some researchers have used crime data to benchmark searches (for instance, Fagan 2002; McMahon et al. 2002). To conduct such an analysis, a researcher should refer to Chapter 10 and our explanation of the use of crime data as a benchmark for stops. Some research teams (for instance, Lovrich et al. 2003; Edwards et al. 2002a, 2002b; Schafer et al. forthcoming) have conducted multivariate analyses to examine searches; we will discuss the strengths and limitations of multivariate analyses in the next chapter. We will also discuss in that chapter how search disparities can be conveyed numerically.

ANALYZING STOP DISPOSITIONS

Does a driver's race/ethnicity have an impact on police behaviors/activities during a vehicle stop? This question was posed at the beginning of this chapter. To address it, jurisdictions can analyze search data. Jurisdictions also can analyze data on stop dispositions (for instance, arrest, citation, warning, no action).

A review of jurisdiction analyses of dispositions is very interesting because of the lack of agreement regarding what results indicate racial bias by police. Some analysts (for instance, the Montgomery County [MD] Police Department 2001) have held that disproportionate representation of minorities among drivers given the most serious dispositions (arrests or citations) is an indication of bias. Other analysts have claimed racial bias is indicated by the disproportionate representation of minorities among those receiving warnings or no disposition (Fagan 2002). Such "low-level" outcomes are not viewed by them as a sign of police benevolence but as evidence that there may have been no legitimate reasons for these stops in the first place. More low-level dispositions for minorities than for Caucasians is seen as evidence of police "fishing" for evidence of crime among minorities.

These varied interpretations of disposition data reflect the challenge researchers face when analyzing this type of data. In their analysis of disposition data, like vehicle stop data, researchers can identify "disparity" in police actions or the lack thereof. They can calculate the percentage of various dispositions across drivers within various racial groups. The results in Table 11.3 for a hypothetical jurisdiction show that minorities are overrepresented among drivers receiving "no disposition." Like the "percent searched" data, disposition data can identify disparity in police actions but not the cause of that disparity.

With regard to legitimate searches of drivers, not all drivers are at equal risk of a search. Similarly, not all stopped drivers are at equal risk of receiving the various dispositions. The ideal benchmark would tell us what the racial/ethnic breakdown should be within each disposition assuming no bias. This

Table 11.3. Stop Dispositions for Caucasians and Minorities

Race	Arrest	Citations	Warning	No Disposition	Total
Minorities	6%	59%	23%	12%	100%
Caucasians	5%	62%	25%	8%	100%

benchmark would take into consideration all of the legitimate factors that can influence the dispositions police choose. In order to fully isolate the cause of disparity, researchers would consider all of these factors unless there were clear evidence that the factors do not vary by racial/ethnic groups.[26]

What are the legitimate factors that might influence police decisions regarding stop dispositions? Lovrich et al. (2003) analyzed Washington State Patrol (WSP) data for the period May 2000 through September 2001. They also reviewed the research on criminal justice decision making (for instance, Black 1980; Matstrofski et al. 2000) in an effort to better understand police decisions on stop dispositions. Lovrich et al. note that this research points to the following key factors influencing an officer's choice of disposition: the seriousness of the offense, the number of offenses committed, the presence of others at the scene, and the demeanor of the subject. They noted that the strongest predictors of disposition behavior among criminal justice officials are number of offenses and seriousness of offenses. Data collected through the WSP vehicle stop form allowed this team to incorporate these factors into their analyses of the disposition data.

Their multistage analyses produced results confirming the importance of these variables in understanding police disposition decisions. First the team conducted the disparity analysis

26 As explained in Chapter 2, the analyst should consider the alternative legitimate factors in the model unless there is clear evidence of no differences across groups.

described above. For each of the 40 subareas (WSP districts), the team looked at the breakdown of dispositions for each racial/ethnic group. These analyses showed that minorities were disproportionately represented among the people getting arrest/citations (versus written warnings or verbal warnings). Specifically, in 31 of 40 of the districts, higher proportions of stopped African Americans were issued citations than stopped Caucasians. This was true for Native Americans as well in 31 of 40 districts and for Hispanics in 39 of 40 districts.

If interpreted by less knowledgeable researchers, these data might have been used to conclude that the Washington State Patrol was practicing racially biased policing. Instead, the team proceeded to the second stage of analysis that took into consideration for each driver the number of violations detected at the time of the stop and the cumulative seriousness of those detected violations. As indicated above, the WSP forms specified the necessary information. Space on the forms allowed officers to report up to eight violations that they observed before or during the course of their interaction with the driver. Based on the information officers recorded regarding the type of offense (for instance, speeding, felony flight), the researchers developed a measure of overall seriousness of combined offenses for each driver. For each identified offense, the violation was coded as either 1 for "serious" or 0 for "other."[27]

They conducted multivariate analyses using as independent variables the number of violations and the seriousness of the violations. Both of these variables had "strong effects" on disposition decisions of officers (Lovrich et al. 2003, 29). When these legitimate, alternative factors were considered in the mul-

---

27 "Serious violations included: felony drugs; misdemeanor drugs; DUI drugs with test; DUI drugs, no test; DUI underage, with test; DUI underage, no test; DUI with test; DUI without test; felony flight, elude; felony warrant; hit and run; insurance-none; license suspension/revocation; misdemeanor warrant; negligent driving, 1st degree; negligent driving, 2nd degree; reckless driving; vehicular homicide; and vehicular assault" (Lovrich et al. 2003, 54).

tivariate models, minorities were no longer disproportionately represented among drivers receiving citations. These results were produced because in many districts African Americans and Hispanics had a higher average number of violations than Caucasians and Asian drivers (p. 52) and higher average seriousness scores (p. 54).

This research highlights how important it is for law enforcement agencies to interpret data responsibly. Researchers should identify and consider in their analysis and/or interpretation of disposition data the nonbias factors that legitimately influence police choices of dispositions. When analyzing stop data, researchers should consider quantity of driving, quality of driving, and location (see Chapter 2). Similarly, when analyzing disposition data, researchers need to consider relevant alternative factors influencing police decisions. These variables are not required for purposes of identifying "disparity," but they are required for isolating the cause of disparity and drawing conclusions regarding possible racial bias by police.

The quantity and seriousness of the violations by the stopped driver appear to be the key variables that influence police disposition decisions, but they are not the only ones.[28] Others might include driver demeanor, prior driving record, and geographic location of the stop. An example will illustrate the importance of stop location. An officer might consider speeding 10 miles per hour over the speed limit in a school zone as a more serious offense than 10 miles per hour over the speed limit on a highway (Schafer, Carter, and Katz-Bannister 2004).

In disposition data analysis, like stop data analysis, the more legitimate factors the researcher can rule out, the more confidence there can be that the disparity in police decisions is

---

[28] Measures by police of number of offenses and seriousness of offenses could themselves be impacted by racial/ethnic bias. See Mosher et al. (2004, 17) for an empirical assessment of whether "members of the Washington State Patrol were deliberately 'piling on' violations or recording more serious violations for minorities in order to justify issuing them citations."

due to bias.[29] The researcher is never going to be able to control for all legitimate, alternative factors. However, a jurisdiction wishing to move beyond the mere measurement of disparity to control for at least some of the key nonbias causes for that disparity can take certain constructive steps. Specifically, it could include on the next iteration of its data collection form the variables used by the WSP team (see next section).[30] We proceed below to guide agencies in how they might analyze, present, and interpret disposition data.

Resources Required

The form that officers fill out should include an item regarding the disposition of the stop. Common options are arrest, ticket/citation, verbal warning, written warning, and no action. Information related to the reasons for stopping the vehicle are relevant to analyzing the dispositions of those stops. Therefore, data collection forms should include a field for "reason for the stop." There is a lot of variation across agencies with regard to the specificity of the "reason" options. The most simplified version for an agency collecting data on all vehicle stops (traffic and investigative stops) might include:

- Moving vehicle code violation,
- Nonmoving vehicle code violation,
- Misdemeanor penal code violation (including suspicion of),
- Felony penal code violation (including suspicion of),
- Other.

---

29 Recall from Chapter 2 that disparity could also be "masked" when key variables are not included in the analysis. A finding of no disparity in an analysis that excludes key alternative variables does not necessarily mean bias does not exist.

30 As with analyzing "who is stopped data," however, it is not practical, arguably impossible, to try to measure all of the factors that might conceivably impact on police disposition decisions.

The following option (Fridell et al. 2001, 126-127) contains more detail:

Vehicle Code Violation:

- Red light/stop sign
- Speed [___miles per hour over the limit]
- Lane violation
- Commercial vehicle
- Following too closely
- Failure to signal
- Other moving violation
- Hazardous equipment
- Seat belt
- Other nonmoving violation

Penal Code Violation:

- Nuisance (related to quality of life)
- Vice
- Property crime
- Violent crime
- Violation of local ordinance
- BOLO/Person wanted
- Suspicious circumstances.

Below we will describe the added value of using these more specific categories. When making decisions regarding form content, however, agencies must balance this added value for researchers against the increased burden on officers who must complete the lengthier form.

The variables included in the Washington State Patrol data that measured quantity and seriousness of violations proved to be valuable. The WSP used an "activity report" to collect the data. There were eight fields that could be filled in by officers that provided information to produce the variables for quantity and seriousness of violations. For each stop, the officers used the first field to list the primary "reason for the stop" (using numerical codes for type of violation) and the remaining fields to list additional violations detected.

Analysis of Data on Reason for a Stop

In the quest to account for the legitimate factors that can influence disposition decisions, researchers should conduct analyses within categories of "reason for a stop." This reduces the variation in violation seriousness at least somewhat. We would expect, for instance, that a disposition for a serious violation, such as 1st degree negligent driving, would be more harsh than a disposition for "driving too closely." Researchers can control for these expected variations in disposition by type of offense by analyzing dispositions by race and ethnicity within categories of "reason for a stop."[31]

If sample size permits, the analyst could conduct separate analyses for each of the various categories included on the jurisdiction's stop form. For instance, Table 11.4 (p. 302) provides hypothetical disposition data for moving violations in Jurisdiction A by race and ethnicity. From the corresponding Figure 11.3, we can see that, relative to the other groups, African Americans are underrepresented among detained persons who receive a citation for moving violations and overrepresented among people who are arrested. We cannot draw conclusions about racial bias based on these data. To do that, we would need to know what the dispositions would be, assuming no bias. It is conceivable (and unknowable from these data) that proportionately more African

[31] This point merits elaboration. The seriousness of the offense is one of the factors that legitimately influences police decisions. For this reason, researchers try to control for or isolate this factor. If researchers examined dispositions for data that included all possible offenses, they would not know if a finding that African Americans received harsher dispositions than Caucasians was due to bias or the possibility that they committed more serious driving violations. Two factors (at least) could be producing the results: violation seriousness or officer bias. Because the types of offenses listed on forms vary by seriousness, researchers are able to analyze the data within those categories to reduce the influence of offense seriousness on results. Instead of doing one analysis of dispositions for all violations combined, researchers are encouraged to look at dispositions across races within offense categories such as speeding violations, red light violations, failure to yield violations, etc.

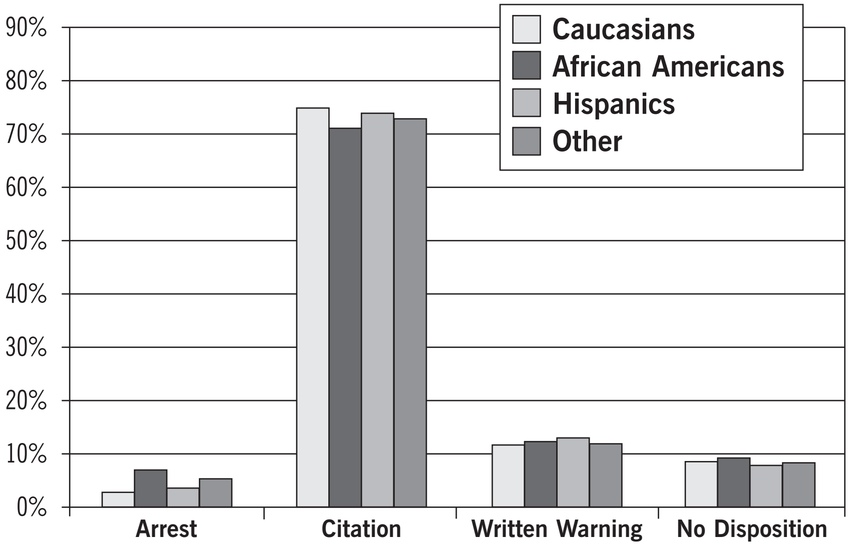

Figure 11.3. Dispositions for Moving Violations, by Race/Ethnicity

Americans than the other groups presented behaviors that legitimately led to the arrest disposition.

What if the data had showed minorities were overrepresented among drivers receiving no disposition? As we noted earlier, overrepresentation of minorities among drivers receiving no disposition has raised legitimate concerns on the part of some observers. They believe such a finding reflects instances where the officers had no legal justification for the stop in the first place. However, from descriptive data such as that presented in Table 11.4, we would not be able to tell whether or not that was the case.[32] In light of the concerns regarding racial profiling in Jurisdiction A, it is possible that officers were letting minorities

32 Smith et al. (2003) note that stops that include searches and result in low-level dispositions (for instance, warnings or no actions) could represent pretext stops by officers who are merely "fishing" for evidence of crime. Of particular concern would be those stops that included nonproductive searches.

Table 11.4. Dispositions for Moving Violations, by Race/Ethnicity, Hypothetical Jurisdiction A

Race	Number Detained	Percent of Detained	Disposition							
			Arrest		Citation		Written Warning		No Disposition	
African Americans	6,405	15.65%	402	6.28%	4,600	71.82%	801	12.51%	602	9.40%
Hispanics	1,700	4.15%	54	3.18%	1,267	74.53%	234	13.76%	145	8.53%
Other Minority	8,182	20.00%	402	4.91%	5,998	73.31%	1,035	12.65%	747	9.13%
Caucasians	24,629	60.19%	623	2.53%	18,772	76.22%	2,997	12.17%	2,237	9.08%
Total	40,916	100.00%	1,481	3.62%	30,637	74.88%	5,067	12.38%	3,731	9.12%

off with verbal warnings to avoid vehicle stop statistics that show harsh dispositions. That said, however, such findings, even if they can't be used to prove or disprove racially biased policing, might lead to discussions and reforms.

We discussed in earlier chapters the potential impact of age, and even gender, on violating behavior. To remove the potential, hidden impact of these variables on driving behavior and thus on dispositions, the researcher could conduct analyses of dispositions within types and within age and gender categories. For instance, Table 11.4 could be completed for young males who were stopped for moving violations. This would provide a comparison of dispositions for young minority males and young Caucasian males who were stopped for moving violations. A researcher hoping to conduct such analyses may find, however, that the number of stops within categories (that is, number of stops involving young men committing moving violations) are too small for reliable analyses, particularly if they are conducted within subareas.

Another constraint associated with interpreting disposition data stems from the ambiguity inherent in interpreting the responses recorded on data collection forms with regard to the event(s) within the stop that led to the disposition. Imagine an officer pulls a driver over for speeding 10 miles per hour over the speed limit. During the course of the stop, the officer asks for consent to search, and it is granted. The officer finds contraband and makes an arrest. A simple analysis of dispositions by "reason for a stop" will show this to be an arrest for speeding 10 miles per hour over the speed limit when, in fact, the arrest was based on the search result. An analyst should assess, based on the form used by the jurisdiction, how such ambiguity might affect his/her analyses and attempt adjustments. For instance, if the jurisdiction's form would produce the misleading link above (an "arrest" for "speeding"), the researcher might choose to conduct separate analyses of dispositions for the stops that did and did not produce positive search results.

Analysis of Data on Levels of Speeding

If the necessary information were available, a researcher could compare dispositions within the very specific "offense seriousness" categories provided by data on miles per hour over the speed limit. Instead of comparing dispositions within the broad "reason for a stop" category of "speeding," a researcher could subdivide this category based on information on the stop form regarding how many miles per hour the person was speeding (see, for instance, Farrell et al. 2004; Dedman and Latour 2003). This produces more refined categories of seriousness of offense for purposes of controlling for this legitimate alternative factor. Such an analysis could produce the equivalent of Table 11.4 for each subcategory of speeding seriousness, such as "5 to 10 miles per hour over the speed limit," "11 to 20 mph over the speed limit," and so forth. Where there are sufficient numbers of stops to support even more refined categories, the variables of age and gender could be included. This type of analysis may be most viable for analyses of state patrol/police data due to the large number of stops for speeding and the fact that speeding stops comprise a large proportion of all stops.[33]

Some research teams (for instance, Edwards et al. 2002a and 2002b; Schafer et al. forthcoming; Crawford 2000; Cox et al. 2001) have conducted multivariate analyses to examine stop dispositions. We will discuss the strengths and limitations of these analyses in the next chapter. We will also discuss in that chapter how disposition disparities can be conveyed numerically.

ANALYZING OTHER ASPECTS OF A STOP

In addition to collecting information on searches and dispositions, some agencies collect other information related to what

---

33 Engel et al. (2004) used information on mph over the speed limit to assess disparity in stop decisions (versus dispositions). The team compared the "average miles per hour over the speed limit" for the speeding stops of minorities and Caucasian drivers to see if minorities were stopped for less severe infractions.

happens after a stop is made. Some jurisdictions, for instance, collect information on the duration of the search or the duration of the entire stop. This might be included on the form as an open-ended question: How many minutes did the search or stop take? The officer would insert the actual number of minutes. Such a question would produce a continuous variable. Alternatively, the form could include response options such as "0–15 minutes," "16–30 minutes," "31–60 minutes," and "61+ minutes."

Agencies might collect information regarding whether the driver (or passengers) were asked to exit the vehicle, whether canines were brought to the scene, and whether firearms were drawn. Although these variables, like the others we've discussed, have limitations in terms of our ability to identify the existence of racial bias, an agency may decide to include one or more of them merely to understand more fully what happens during traffic stops in their jurisdiction.

The general analysis concepts presented above, indeed throughout this book, apply to these and any other variables. The first question for the researcher to ask is as follows: what factor, other than racial bias, might account for different decisions/actions by police? For duration of the stop, the analyst would want to consider the legitimate factors that might make a stop longer or shorter. Some of these factors might be measured in the form, others will not be. An example of a factor that will lengthen a stop and is likely to be included on the data form is search activity; stops involving searches are likely to be longer than those that do not. An arrest disposition is another example of a factor that would lengthen a stop and be available on the form. With such information, the analyst could assess disparity across stop duration controlling for the occurrence of an arrest and/or search. The researcher might compare length of stop across Caucasians and minorities for each of the following categories (1) stops that do not involve either a search or arrest, (2) stops that involve a search, (3) stops that involve an arrest, and (4) stops that involve both a search and arrest.

CONCLUSION

There has been considerable development recently in the collective thinking about how to analyze search data. A researcher can calculate and compare "percent searched" for racial/ethnic groups to indicate whether disparity exists, but cannot with these data draw conclusions about the existence or lack of racial bias. Similarly, hit rates for all types of searches can provide an indication of whether disparity exists. Hit rates that meet the assumptions of the outcome test can indicate the existence of unjustified disparate impact. The searches that meet the assumptions are those where the officers' decision to search is based on the probability of finding contraband/evidence. Different hit rates for minorities and Caucasians for these evidence-based searches should lead an agency to consider additional assessments of searches or reform measures. Unfortunately, consent searches—which are high-discretion searches for officers and therefore vulnerable to manifestations of bias—usually cannot be analyzed with the outcome test.

Disposition data analysis has also been an interesting area of study with vastly different interpretations by analysts of similar results. The team analyzing the data for the Washington State Patrol showed the potential impact of two nonbias factors—the number of offenses and the seriousness of offenses by stopped drivers—on the choice of disposition. Although most agencies do not have information regarding these variables, they can identify disparities in disposition decisions by comparing officers' decisions within categories of types of stops as defined by "reason for the stop." Such results would be reported with caution. From the data, agencies can highlight areas of disparity and areas for potential concern, but they cannot draw conclusions regarding bias by police.

The analysis of poststop data is complicated, and most methods can indicate only whether disparity exists, not the cause. Despite these constraints, researchers should analyze poststop data and report to the law enforcement agency and other stakeholders comprehensive information regarding what happens after stops are made. These poststop activities are vulnerable to racial bias by police, and they could have great negative consequences for the driver subject to them. It is important for police executives to know what is happening during vehicle stops since these incidents comprise the most frequent interaction between police and citizens. As we discuss in Chapter 13, a finding of disparity, even if the cause of the disparity cannot be identified, can provide impetus for constructive changes in law enforcement policies or practices.

XII Drawing Conclusions from the Results

Previous chapters have explained ways in which data on vehicle stops by police and data on poststop activity by police (for example, searches and dispositions) can be analyzed. Jurisdictions are trying to determine whether there is a cause-and-effect relationship between a driver's race/ethnicity and police behavior.

Chapter 2 set forth the benchmarking challenge. Researchers take stop data collected by police and attempt to develop a comparison group to produce a "benchmark" against which to measure this data. Benchmarking is a comparison of the racial/ethnic profile of the people identified in the police-citizen contact data and the racial/ethnic profile of a "benchmark population." This population might be composed of residents of the jurisdiction with access to vehicles (Chapter 5, "Benchmarking with Adjusted Census Data"), drivers with a license (Chapter 6, "Benchmarking with DMV Data"), drivers identified by red light cameras, radar, or air patrols (Chapter 7, "Benchmarking with Data from 'Blind' Enforcement Mechanisms"), drivers stopped by "matched" officers or groups of officers (Chapter 8 on internal benchmarking), drivers observed on the road by researchers (Chapter 9, "Observation Benchmarking"), or drivers identified through other benchmarking methods (Chapter 10).

In conveying the information produced by the various benchmarking methods, we produced figures similar to Figure 12.1. This figure compares minorities and Caucasians in terms of their representation among people stopped and among the benchmark population. Minorities are overrepresented among drivers stopped relative to their representation in the benchmark population. They represent 19.06 percent of the stopped drivers and 15.60 percent of the benchmark population. Figure 12.1 indicates that disparity exists. As noted in Chapter 2, it is not difficult to measure whether there is disparity between racial/ethnic groups in terms of stops made by police; the difficulty comes in identifying the causes for disparity. Previous chapters have described legitimate causes for disparity and how researchers, using each benchmarking method, can attempt to rule them out before making any claims that the identified dis-

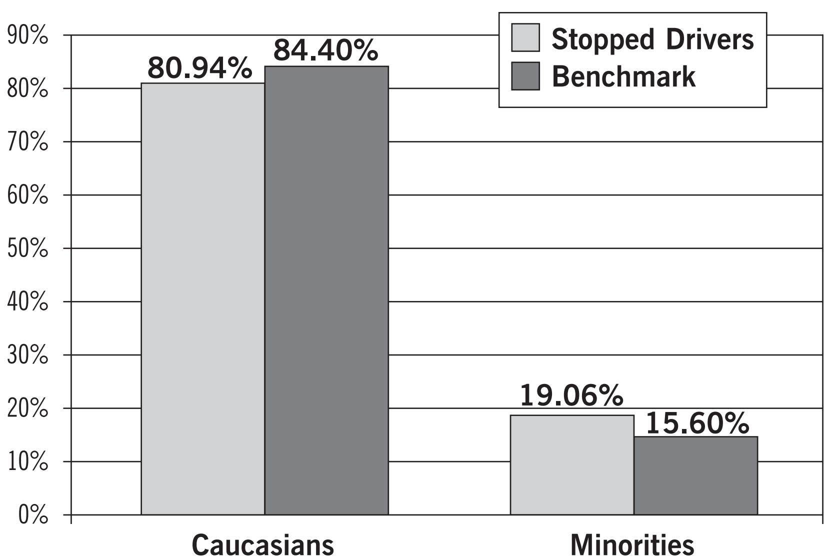

Figure 12.1. Disparity between Drivers Stopped by Police in Hypothetical Area A and the Benchmark Population for Area A, by Two Racial/Ethnic Groups

parity is the likely result of police bias. The focus of this chapter is not those causes of disparity but rather how measures of disparity can be conveyed and interpreted.

We begin the chapter by reporting four ways that disparity—such as that shown in Figure 12.1—can be conveyed: through absolute percentage differences, relative percentage differences, disparity indexes, and ratios of disparity. We then explain how these four measures of disparity can be calculated using stop, search, and disposition data. After describing the various measures of disparity, we discuss the factors that a researcher should consider when deciding how many of these measures to report and which ones. Two additional tools for assessing and conveying disparity—contingency analysis and multivariate analyses—are described along with tips for their use and caveats.

We then return to the crux of the matter: when does disparity between the racial/ethnic profile of stopped drivers (the numerator data) and the racial/ethnic profile of the benchmark population (the denominator data, see Chapter 4) equate to bias? There is no simple answer to this question, but we will present practical suggestions based on the work of the social scientists analyzing vehicle stop data. Some advocate setting a cut-off point whereby disparity levels above it indicate racial bias and disparity levels below it indicate none; others believe it is impossible, and therefore inappropriate, to set a cut-off point. We evaluate these opinions and explain useful tools that can help researchers interpret data that indicate disparity. We also describe how researchers can conduct or facilitate a "qualitative review of quantitative data."

The contents of this chapter will generate frustration in many researchers who are under pressure from the consumers of their reports (for instance, police chiefs, community leaders, journalists, politicians, and other stakeholders) to provide definitive answers regarding whether or not policing in their jurisdiction is characterized by racial bias. A theme of this book is that we can measure disparity easily but identifying the cause of disparity presents a challenge. That theme continues through this chapter. No calculations of measures of disparity—however advanced—will themselves overcome this challenge. Researchers who adhere to sound principles of social science will recognize the conclusions that can and cannot be drawn from the results of benchmarking analysis.

CALCULATING MEASURES OF DISPARITY FOR DATA ON "WHO IS STOPPED"

Figure 12.1 showed disparity between stopped drivers and the benchmark population for Area A of a hypothetical jurisdiction. To simplify this initial explanation, we have separated citizens into just two groups: minorities and Caucasians. Table 12.1 presents four ways to measure and convey the disparity indicated in Figure 12.1. Column A presents the number of stops of minorities and Caucasians across the reference period (for instance, one year). Researchers who are calculating measures of disparity should include in their tables the number of stops so that the discerning reader can assess whether this number is sufficient to produce reliable results.[1]

Column B presents the percentage of the stops by police that were of minorities and of Caucasians (summing to 100 percent). Thus, for instance, the percentage of stops that were of minori-

---

1 Analyses with small numbers of stops are less reliable than those with larger numbers of stops. In some circumstances, the researcher can achieve reliable numbers by combining categories in the analysis. For instance, a researcher may combine categories of racial groups. Table 12.1 shows all minorities combined into one group; in other circumstances the researcher may be able to retain a particular racial/ethnic group (for instance, African Americans or Hispanics) but may need to combine the remaining racial/ethnic groups into an "other" category. While combining racial/ethnic groups can produce more reliable analyses, there are drawbacks. Combining groups reduces the specificity of results, which makes the results less useful for policy makers. Stakeholders also may object to combining distinct minority groups into one category. If, for these reasons, a researcher decides not to combine small racial/ethnic groups into one, the researcher should provide caveats with all results. For example, this statement could be made: "Because of the small size of this group, the results are not necessarily reliable and/or generalizable."

Table 12.1. Four Disparity Measures to Describe Stops in Hypothetical Area A, for Two Racial/Ethnic Groups

	A	B	C	D	E	F	G
	Number of Stops	**Percent of Stops**	**Percent of Benchmark**	**Absolute % Difference**	**Relative % Difference**	**Disparity Index**	**Ratio of Disparity**
Equation		[A(m or c)/t] x 100		B-C	[(B-C)/C] x 100	B/C	F(m)/F(c)
Minorities (m)	15,492	19.06%	15.60%	3.46%	22.18%	1.22	1.27
Caucasians (c)	65,789	80.94%	84.40%	-3.46%	-4.10%	0.96	
Total (t)	81,281	100.00%	100.00%				

Note: These data produced the summary results presented in Figure 12.1.

ties is 19.06 [(15,492/81,281) x (100)]. Column C presents from Figure 12.1 the percentage of minorities and Caucasians in the benchmark population. If the jurisdiction were implementing benchmarking with adjusted census data (for instance, adjusted for access to vehicles), Column C would indicate that 15.60 percent of the jurisdiction's residential population with access to vehicles were minorities, 84.40 percent were Caucasians. If the benchmark represented people observed violating speeding laws (as opposed to jurisdiction residents), Column C would indicate that 15.60 percent of the people speeding on the jurisdiction's roads were minorities and 84.40 percent were Caucasians.

Column D shows the first of four ways to convey the disparity indicated in Columns B and C. Column D shows the absolute differences in percentages between those stopped by police and the benchmark population. Column C (representation of the group among the benchmark population) is subtracted from Column B (representation of the group among the drivers stopped by police). For the minority group, the absolute percentage difference is 3.46 percent (19.06% - 15.60%). This result can be conveyed in the following language: "there are 3.46 percent more minorities among the people who are stopped than are represented in the benchmark group."

A second way that researchers can convey disparity is through relative differences in percentages between those stopped by police and the benchmark population. For the minority group in Table 12.1, the relative percentage difference is 22.18 percent or [(19.06-15.60)/15.60] x 100. In other words, 19.06 percent is 22.18 percent greater than 15.60 percent. The language chosen to explain the relative percentage difference in this example could be as follows: "there are 22.18 percent more minorities among the people who are stopped than are represented in the benchmark group." Or, "minorities are over-represented among people stopped by 22.18 percent relative to their representation among the benchmark group. Similarly, whites are under-represented among people stopped by 4.10 percent relative to their representation among the benchmark group." This wording is the same as that which can be used to describe absolute (as opposed to relative) differences in percentages. There is no particular language for conveying the results that distinguishes the figures that are absolute percentage differences and relative percentage differences. Researchers should convey the meaning of the disparity by describing in the report the equation used: either B-C or [B-C/C] x 100 (see Table 12.1).

A third way to convey disparity is using a "disparity index." For the minority group in Table 12.1, the disparity index is 1.22, which is calculated by dividing Column B (group percentage among drivers stopped) by Column C (group percentage among benchmark population). A value of 1 would indicate no disparity; that value would be obtained in our example if 19.06 percent of the stops were of minorities, and minorities comprised 19.06 percent of the benchmark population. A value greater than 1 indicates over-representation among drivers stopped relative to the benchmark, and a value less than 1 indicates under-representation among drivers stopped relative to the benchmark. The results in Table 12.1 indicate an over-representation of minorities among stops relative to their representation in the benchmarked group.[2]

A "ratio of disparity" is the fourth way a finding of disparity can be conveyed.[3] The disparity index for one group is divided by the disparity index for another group. The group in the denominator is the "reference group" to which the other group is compared. In our example, we use the disparity index to gauge how minorities (the numerator in the equation) fare relative to Caucasians (the denominator in the equation).

---

2 Consistent with our caveat that small sample sizes produce unreliable results, note that all of these measures are unstable when sample sizes are small.

3 Harris (1999) and Lamberth (2001) refer to this calculation as producing an "odds ratio." We prefer "ratio of disparity" to reflect the actual equation used to produce it and to avoid reference to "odds," which implies the formuia produces a measure of probabilities, which it does not.

For the minority group in Table 12.1, the ratio of disparity is 1.27 (1.22/0.96). The disparity index for minorities is divided by the disparity index for Caucasians to produce a single number. A number greater than 1 indicates over-representation, and a number less than 1 indicates under-representation. Researchers could explain the ratio of disparity shown in Table 12.1 in any of the following ways:

- "Minorities are stopped 1.27 times more than Caucasians."
- "If you are a minority, you are 1.27 times more likely to be stopped by police than if you are Caucasian."
- "For every Caucasian stopped, 1.27 minorities are stopped."

Table 12.2 shows how to calculate ratios of disparity when there are more than two racial/ethnic groups. Because Hispanics comprised 8.24 percent of the stops and a very similar percent of the benchmark population (8.20 percent), the disparity index for Hispanics is 1.00 (8.24/8.20), indicating no disparity. The disparity indexes for African Americans and Caucasians show over-representation of African Americans relative to the benchmark (1.46) and under-representation of Caucasians (0.96). Recall that to produce the ratio of disparity for the two groups in Table 12.1, we divided the disparity index for minorities by the disparity index for Caucasians (1.22/0.96 = 1.27). To calculate the ratio of disparity with three racial/ethnic groups, researchers again must identify which of the three groups is the "reference group." The disparity index for this chosen reference group becomes the denominator for the ratio of disparity calculations for the other two.

We argue that the relevant reference group in any calculation of a ratio of disparity for vehicle stop analysis is the Caucasian group. This is because the main question we are trying to answer is as follows: "Are minority residents treated differently from Caucasian residents because of their

Table 12.2. Disparity Indexes and Ratios of Disparity to Describe Stops in Hypothetical Area A, for Three Racial/Ethnic Groups

	A	B	C	F	G	
	Number of Stops	Percent of Stops	Percent of Benchmark	Disparity Index (B/C)	Ratio of Disparity	
					Formula	Result
African Americans (a)	8,798	10.82%	7.40%	1.46	F(a)/F(w)	1.53
Hispanics (h)	6,694	8.24%	8.20%	1.00	F(h)/F(w)	1.05
Caucasians (w)	65,789	80.94%	84.40%	0.96		
	81,281	100.00%	100.00%			

racial/ethnic status?"[4] In Table 12.2, the disparity ratio for African Americans is 1.53: the disparity index for African Americans (1.46) divided by the disparity index for Caucasians (0.96). Similarly, to find the disparity ratio for Hispanics, divide the disparity index for Hispanics (1.0) by the disparity index for Caucasians (0.96) to get 1.05. These results in Table 12.2 can be conveyed in any of the three ways described above, including: "African Americans are stopped 1.53 times more than Caucasians. Hispanics are stopped 1.05 times more than Caucasians."

CALCULATING MEASURES OF DISPARITY FOR DATA ON "WHO IS SEARCHED"

Percentage differences (absolute and relative), disparity indexes, and ratios of disparity can be used to describe search data as well as stop data (Tables 12.3 and 12.4). In order to calculate absolute and relative percentage differences for search data, researchers begin the same way they would to calculate percentage differences for stop data: they begin with the number of stops for each group and the percentage of stops for each group (Columns A and B in Table 12.3). For the African American group, the percentage of stops (10.82) is calculated by taking the number of stops of African Americans (8,798), dividing it by the total number of stops (81,281), and multiplying by 100.

Column C shows the number of searches for each racial/ethnic group. Column D shows the percentage of all searches that

---

[4] Others argue that each racial/ethnic group should be compared to all drivers not in that particular racial/ethnic group. Researchers then would compare African Americans to all drivers who are not African Americans instead of to the Caucasian subgroup we used. Similarly, the reference group for Hispanics would be all drivers who are not Hispanics. If the researcher had three racial/ethnic groups (for instance, African Americans, Hispanics, Caucasians), the researcher would produce the ratios of disparity by dividing the African American disparity index by the disparity index for all other drivers combined into one group, and then by dividing the Hispanic disparity index by the disparity index for all other drivers combined into one group. If the resulting ratio of disparity for African Americans was 1.37, the interpretation would be "If you are an African American, you are 1.37 times more likely to be stopped by police than if you are not African American."

Table 12.3. Percentage Differences to Convey Search Data in Hypothetical Area A, for Three Racial/Ethnic Groups

	A	B	C	D	E	F	G
	Number of Stops	Percentage of Stops	Number of Searches	Percentage of Searches	Percent Stops Resulting in Searches	Absolute % Difference	Relative % Difference
Equation		A(a,h or c)/A(t) x 100		C(a,h or c)/C(t) x 100	C/A x 100	E(a or h)-E(c)	E(a or h)-E(c)/E(c)
African Americans (a)	8,798	10.82	1,549	22.36	17.61	10.61	151.53
Hispanics (h)	6,694	8.24	775	11.18	11.58	4.58	65.40
Caucasians (c)	65,789	80.94	4,605	66.46	7.00		
Total (t)	81,281	100.00	6,929	100.00			

were of African Americans, Hispanics, and Caucasians; the data indicate that 22.36 percent of all searches were of African Americans, 11.18 percent of all searches were of Hispanics, and 66.46 percent of all searches were of Caucasians. Column D provides information needed to develop the disparity index. Column E gives the percentage of stops within each racial/ethnic group that resulted in a search. For the African American group, 17.61 percent of all African American stops resulted in searches. The number of African American searches (1,549) is divided by the number of African American stops (8,798) and the result is multiplied by 100.

For the stop data, we calculated the absolute percentage difference by subtracting the representation of the group among the benchmark population from the representation of the group among drivers stopped by police. For search data, we do not have such a clear benchmark to use for comparison purposes. A researcher could convey descriptive information using an absolute percentage difference by subtracting the representation of the group among drivers stopped by police from the representation of the group among drivers searched. Alternatively, the researcher could, as we did in Table 12.3, convey how African Americans fared compared to Caucasians. The absolute difference between the percentage of African Americans searched and the percentage of Caucasians searched is 10.61 percent (17.61% - 7.00%), which is shown in Column F, row 1. A researcher could express this measure of disparity as follows: "10.61 percent more stopped African Americans than stopped Caucasians were searched." The relative difference between the percentage of African Americans searched and the percentage of Caucasians searched is 151.53 percent (the percent of African American stops resulting in searches, minus the percent of Caucasian stops resulting in searches, divided by the percent of Caucasian stops resulting in searches). The results can be expressed in this language: "151.53 percent more stopped African Americans are searched than are stopped Caucasians. Similarly, 65.40 percent more stopped Hispanics are searched than are stopped Caucasians."

Table 12.4. Disparity Indexes, Ratios of Disparity, and Ratios of Stops per Search for Search Data in Hypothetical Area A, for Three Racial/Ethnic Groups

	A	B	C	D	E	F	G
	Number of Stops	**Percentage of Stops**	**Number of Searches**	**Percentage of Searches**	**Disparity Index**	**Ratio of Disparity**	**Ratio of Stops Per Search**
Equation		A(a,h or c)/A(t) x 100		C(a,h or c)/C(t) x 100	D/B	E(a or h)/E(c)	A/C
African Americans (a)	8,798	10.82	1,549	22.36	2.07	2.52	5.68
Hispanics (h)	6,694	8.24	775	11.18	1.36	1.65	8.64
Caucasians (c)	65,789	80.94	4,605	66.46	0.82		14.29
Total (t)	81,281	100.00	6,929	100.00			

Using information presented in Table 12.3, Table 12.4 shows how researchers can develop a disparity index for three racial/ethnic groups by using search data. Divide Column D for each group (representation among searches) by Column B for each group (representation among stops). Again, a value greater than 1 indicates over-representation of drivers searched relative to drivers stopped, and a value less than 1 indicates under-representation. In this example, African Americans and Hispanics are both over-represented among those searched relative to their representation among those stopped (2.07 and 1.36, respectively). Caucasians are under-represented (0.82).

To calculate the ratio of disparity, we divide the disparity indexes for each of the two minority groups (Column E) by the disparity index for Caucasians to produce ratios of disparity for African Americans (2.52) and for Hispanics (1.65). These ratios of disparity could be conveyed in a report to the jurisdiction in this language: "For every stopped Caucasian searched, 2.52 stopped African Americans and 1.65 stopped Hispanics are searched." The report could convey the same results this way: "For those who are stopped, if you are African American, you are 2.52 times more likely to be searched than if you are Caucasian; if you are Hispanic, you are 1.65 times more likely to be searched than if you are Caucasian." Another possible wording is as follows: "Stopped African Americans are searched 2.52 times more than stopped Caucasians. Stopped Hispanics are searched 1.65 times more than stopped Caucasians."

The same search data indicating disparity can be presented one final way: in terms of the ratio of stops per search. By dividing Column A by Column C, the researcher finds that there is one search for every 5.68 stops of African Americans, one search for every 8.64 stops of Hispanics, and one search for every 14.29 stops of Caucasians.[5]

---

[5] Table 12.4 calculates measures of disparity for all types of searches combined. A researcher could create similar individual tables for subsets of searches (for instance, consent searches, warrant searches, evidence-based searches).

CALCULATING DISPARITY IN SEARCH HIT RATES

As reported in Chapter 11, hit rates are the percentage of searches that result in a hit; if 4 of the 80 searches of African Americans produced contraband or other evidence, the "hit rate" would be 5 percent (4/80 x 100). For all types of searches, hit rates can be used to measure disparity. For any type of search, researchers can determine whether searches are more productive for one racial/ethnic group than another. For the types of searches that meet the assumptions of the outcome test, researchers can use hit rates to determine if there is unjustified disparity.

As explained in the previous paragraph, hit rate data can be presented very simply: the percentage of searches that result in hits. In Table 12.5 this simple calculation is presented in Column C. Of the 2,324 searches of minorities in Area B, 220 produced contraband/evidence ("hits"); of the 4,605 searches of Caucasians, 691 resulted in hits. This produces hit rates of 9.47 percent for minorities and 15.01 percent for Caucasians.

These results indicate that the searches of minorities are less productive than the searches of Caucasians. If the searches included in Table 12.5 were limited to ones that met the assumptions of the outcome test explained in Chapter 11 (evidence-based searches), then further exploration by the police department and even intervention might be warranted. But no conclusions regarding the cause or causes of the difference in hit rates of racial/ethnic groups can be drawn if the data include searches that do not meet the assumptions of the outcome test. The researcher could note the difference in search productivity across the two groups, but he or she could not claim that this difference was caused by racial bias. The alternative hypotheses to the bias hypothesis have not been addressed.

We have begun with a simple calculation. Hit rate data can also be presented in a more complex way. This data, like stop data, can be used to calculate a disparity index (Column F in Table 12.5) and a ratio of disparity (Column G) for racial/ethnic groups.

Table 12.5. Disparity Measures for Search Hit Rate Data in Hypothetical Area B, for Two Racial/Ethnic Groups

	A	B	C	D	E	F	G	H
	Number of Searches	Number of Hits	Percent Hits	Percent of Searches	Percent of Hits	Disparity Index	Ratio of Disparity	Ratio of Searches Per Hits
Equation			(B/A) x100	A(m or c)/A(t)	B(m or c)/B(t)	E/D	F(m)/F(c)	A/B
Minorities (h)	2,324	220	9.47	33.54	24.15	0.72	0.63	10.56
Caucasians (c)	4,605	691	15.01	66.46	75.85	1.14		6.66
Total (t)	6,929	911		100.00	100.00			

The disparity indexes in Table 12.5 confirm that searches of minorities are less productive than searches of Caucasians. From those indexes, the researcher can produce the ratio of disparity. In our example, "minority searches are only 0.63 times (approximately "two thirds") as productive as Caucasian searches." The researcher also can say this about the jurisdiction: "there are 10.56 searches of minorities for every one search that results in a hit. In contrast, there are 6.66 searches of Caucasians for every one search that results in a hit." Later in the chapter we will discuss what conclusions concerning racial bias by police can and cannot be drawn from such measures of disparity.

CALCULATING DISPARITY FOR DISPOSITION DATA

In addition to stop data and search data, researchers analyze data on the disposition chosen by police after stopping a driver. As explained in Chapter 11, possible dispositions include arrest, citation, written warning, and no action. This data can indicate disparity in the dispositions given drivers in different racial/ethnic groups. A disparity index and ratio of disparity are shown in Table 12.6, which presents results for two of the four dispositions listed in Table 11.4. The disparity indexes for the arrest data (Column C in Table 12.6) show that African Americans and Other Minorities are over-represented among people arrested relative to their representation among people stopped. Hispanics and Caucasians are under-represented. This same information is conveyed with the ratio of disparity in Column D, with Caucasians as the reference group. Here we see that "Stopped African Americans are arrested 2.47 times more than are stopped Caucasians. Stopped Hispanics are arrested 1.26 times more than are stopped Caucasians. Stopped Other Minorities are arrested 1.94 times more than are stopped Caucasians." In Table 12.6, Other Minorities and Caucasians are provided with "no action" dispositions proportionate to their representation in the stopped population. This result reflects, for instance, the match between the Caucasian representation among people stopped (60.19 percent) and the

Table 12.6. Disparity Measures for Two Dispositions, Arrest and No Action, in a Hypothetical Jurisdiction, for Four Racial/Ethnic Groups

		Arrest Disposition			No Action Disposition		
	A	B	C	D	E	F	G
	Percent of Stopped	Percent of Arrested	Disparity Index	Ratio of Disparity	Percent of No Action	Disparity Index	Ratio of Disparity
Equation			B/A	C(a,h or o)/C(c)		E/A	F(a,h or o)/F(c)
African Americans (a)	15.65%	27.14%	1.73	2.47	16.14%	1.03	1.04
Hispanics (h)	4.15%	3.64%	0.88	1.26	3.89%	0.94	0.94
Other Minorities (o)	20.00%	27.14%	1.36	1.94	20.02%	1.00	1.00
Caucasians (c)	60.19%	42.07%	0.70		59.96%	1.00	
Total (t)	99.99%	99.99%			100.01%		

Caucasian representation among people provided with a "no action" disposition (59.96 percent). African Americans who are stopped are slightly more likely to be given a "no action" disposition than Caucasians who are stopped (Column G).

THE CHALLENGE OF SELECTING MEASURES OF DISPARITY

So far this chapter has explained four different ways that researchers can convey disparity: absolute percentage difference, relative percentage difference, disparity index, and ratio of disparity. For stop, search, and disposition data, the chapter has described not only the formulas for calculating these measures of disparity but also the language that researchers can use to explain to the public what the mathematical measures mean. We turn now to a new question: Which measure or measures of disparity should researchers select to present their data?

Social scientists analyzing vehicle stop data have differences of opinion regarding whether researchers should report multiple measures of disparity or just one. Those who advocate the selection and reporting of a single measure (for instance, the disparity index) point out that multiple measures could confuse those who read the law enforcement agency's report—policy makers, residents, and other stakeholders. The use of multiple measures might lead the various stakeholders with different concerns or agendas to pick and choose the figures in the report that confirm their views or preconceived expectations regarding the results.

Other social scientists favor reporting two, three, or even all four of the measures of disparity. They claim it is better to provide report consumers with more information, not less, including information on how various measures can produce different results in different circumstances.

This fact—that different measures produce different results—is relevant to the researcher who chooses to use one measure of disparity and the researcher who chooses to use all four. As we highlight in the next section, care must be exercised when interpreting any measure of disparity. When a researcher is not dealing with very high or very low percentages of minorities in the population of stopped drivers or in the benchmark population, then the selection of one measure over another does not have strong ramifications for the results. On the other hand, when a researcher is dealing with high or low percentages of minorities, the selection of one measure over another could produce a very different interpretation of results, as we will now explain.

Different Measures of Disparity: Different Interpretations

It would not matter which measure or measures were selected to convey results if all the measures tracked each other in a linear fashion under all circumstances. However, this is not the case. The four measures can convey very different results. Because of these differences, the conclusions a researcher draws based on one measure could be very different from the conclusions the researcher would draw if he or she had selected another measure.

Table 12.7 illustrates this point. It shows four measures of disparity for three hypothetical police departments: A, B, and C. Which department has the most disparity?[6] Well, the answer depends on the measure of disparity we consider. Looking at the absolute percentage difference, we see that Department C has the most disparity. African Americans are over-represented in the stop data relative to the benchmark data by 13.0 percent. Looking at the other three measures of disparity, however, we see that Department B has the most disparity. Although Department B has an absolute percentage difference of only 0.7, it has a relative percentage difference of 117. The disparity index and ratio of disparity for Department B are both 2.2. Department A has the second highest disparity when disparity is calculated as the relative percentage difference (56 percent)

[6] We are referring to disparity between the stopped driver population and the benchmark population in terms of African American representation.

Table 12.7. Various Measures of Disparity for Hypothetical Departments A, B, and C

Department	Representation of African Americans Among Stops	Representation of African Americans Among Benchmark	Absolute % Difference	Relative % Difference	Disparity Index	Ratio of Disparity
A	14.0%	9.0%	5.0%	56.0%	1.6	1.6
B	1.3%	0.6%	0.7%	117.0%	2.2	2.2
C	67.0%	54.0%	13.0%	24.0%	1.2	1.7

Source: Farrell 2004

or disparity index (1.6); Department C has the second highest disparity (1.7) when calculated as ratios of disparity. Clearly, the measure chosen makes a difference in terms of the level of disparity indicated. As a result, before a researcher draws and reports conclusions about disparity in a jurisdiction, he or she should consider what the picture would look like if another measure of disparity had been selected to convey the results.

When the percentage of minorities (or of Caucasians) in both the stopped driver population and the benchmark population is low, the variation between two of the measures of disparity is extreme.[7] Those two measures are the absolute percentage difference and the relative percentage difference. We can see this in Table 12.7 for Department B. Minorities represent only 1.3 percent of the persons stopped and only 0.6 percent of the benchmark population; the absolute percentage difference is tiny (0.7 percent), but the relative percentage difference is large (117 percent).

This extreme variation is even more evident in Table 12.8. In order to highlight the effects of low levels of minorities in the stop and benchmark populations on the four measures of disparity, we arbitrarily set the absolute percentage difference at 2 percent for thirty-five hypothetical departments. In other words, for the sake of example, we say that the absolute percentage difference between the minority representation in the stop data and the minority representation in the benchmark data for all thirty-five jurisdictions is 2 percent. This measure of disparity is fixed. But the departments vary from 2 percent to 100 percent in terms of the minority representation in stops. The top row shows that minorities represent 2 percent of stopped drivers and 0 percent of the benchmark population (and therefore Caucasians represent 98 percent of those stopped and 100 percent of those in the benchmark). Each row in suc-

---

7 Here we focus on the situation when the percentage of minorities is low in the stop and/or benchmark populations. The same problems would occur if Caucasians were the group with low percentage representation.

Table 12.8 Disparity Measures for Multiple Departments When Absolute Percentage Difference is Set at Two

	Percent of Stops		Percent of Benchmark		Percentage Difference		Disparity Index		Ratio of
Dept.	Caucasians	Minorities	Caucasians	Minorities	Absolute	Relative	Minority	Caucasian	Disparity
1	98	2	100	0	2.0	NA*	NA*	0.98	1.02
2	97	3	99	1	2.0	200.00%	3.00	0.98	3.06
3	96	4	98	2	2.0	100.00%	2.00	0.98	2.04
4	95	5	97	3	2.0	66.67%	1.67	0.98	1.70
5	94	6	96	4	2.0	50.00%	1.50	0.98	1.53
6	93	7	95	5	2.0	40.00%	1.40	0.98	1.43
7	92	8	94	6	2.0	33.33%	1.33	0.98	1.36
8	91	9	93	7	2.0	28.57%	1.29	0.98	1.31
9	90	10	92	8	2.0	25.00%	1.25	0.98	1.28
10	89	11	91	9	2.0	22.22%	1.22	0.98	1.25
11	88	12	90	10	2.0	20.00%	1.20	0.98	1.23
12	83	17	85	15	2.0	13.33%	1.13	0.98	1.16
13	78	22	80	20	2.0	10.00%	1.10	0.98	1.13
14	73	27	75	25	2.0	8.00%	1.08	0.97	1.11
15	68	32	70	30	2.0	6.67%	1.07	0.97	1.10
16	63	37	65	35	2.0	5.71%	1.06	0.97	1.09
17	58	42	60	40	2.0	5.00%	1.05	0.97	1.09
18	53	47	55	45	2.0	4.44%	1.04	0.96	1.08
19	48	52	50	50	2.0	4.00%	1.04	0.96	1.08
20	43	57	45	55	2.0	3.64%	1.04	0.96	1.08
21	38	62	40	60	2.0	3.33%	1.03	0.95	1.09
22	33	67	35	65	2.0	3.08%	1.03	0.94	1.09
23	28	72	30	70	2.0	2.86%	1.03	0.93	1.10
24	23	77	25	75	2.0	2.67%	1.03	0.92	1.12
25	18	82	20	80	2.0	2.50%	1.03	0.90	1.14
26	13	87	15	85	2.0	2.35%	1.02	0.87	1.18
27	8	92	10	90	2.0	2.22%	1.02	0.80	1.28
28	7	93	9	91	2.0	2.20%	1.02	0.78	1.31
29	6	94	8	92	2.0	2.17%	1.02	0.75	1.36
30	5	95	7	93	2.0	2.15%	1.02	0.71	1.43
31	4	96	6	94	2.0	2.13%	1.02	0.67	1.53
32	3	97	5	95	2.0	2.11%	1.02	0.60	1.70
33	2	98	4	96	2.0	2.08%	1.02	0.50	2.04
34	1	99	3	97	2.0	2.06%	1.02	0.33	3.06
35	0	100	2	98	2.0	2.04%	1.02	NA*	NA*

*Not applicable because formula places a zero in the denominator of the equation.

Table 12.9. Measures of Disparity for Multiple Departments When Absolute Percentage Difference is Set at 15

	Percent of Stops		Percent of Benchmark		Percentage Difference		Disparity Index		Ratio of
Dept.	Caucasians	Minorities	Caucasians	Minorities	Absolute	Relative	Minority	Caucasian	Disparity
1	100	0			Not poss				
2	95	5			Not poss				
3	90	10			Not poss				
4	85	15	100	0	15.0	NA*	NA*	0.85	NA*
5	84.5	15.5	99.5	0.5	15.0	3000.00%	31.00	0.85	36.50
6	84	16	99	1	15.0	1500.00%	16.00	0.85	18.86
7	83	17	98	2	15.0	750.00%	8.50	0.85	10.04
8	82	18	97	3	15.0	500.00%	6.00	0.85	7.10
9	81	19	96	4	15.0	375.00%	4.75	0.84	5.63
10	80	20	95	5	15.0	300.00%	4.00	0.84	4.75
11	79	21	94	6	15.0	250.00%	3.50	0.84	4.16
12	78	22	93	7	15.0	214.29%	3.14	0.84	3.75
13	77	23	92	8	15.0	187.50%	2.88	0.84	3.44
14	76	24	91	9	15.0	166.67%	2.67	0.84	3.19
15	75	25	90	10	15.0	150.00%	2.50	0.83	3.00
16	70	30	85	15	15.0	100.00%	2.00	0.82	2.43
17	65	35	80	20	15.0	75.00%	1.75	0.81	2.15
18	60	40	75	25	15.0	60.00%	1.60	0.80	2.00
19	55	45	70	30	15.0	50.00%	1.50	0.79	1.91
20	50	50	65	35	15.0	42.86%	1.43	0.77	1.86
21	45	55	60	40	15.0	37.50%	1.38	0.75	1.83
22	40	60	55	45	15.0	33.33%	1.33	0.73	1.83
23	35	65	50	50	15.0	30.00%	1.30	0.70	1.86
24	30	70	45	55	15.0	27.27%	1.27	0.67	1.91
25	25	75	40	60	15.0	25.00%	1.25	0.63	2.00
26	20	80	35	65	15.0	23.08%	1.23	0.57	2.15
27	15	85	30	70	15.0	21.43%	1.21	0.50	2.43
28	10	90	25	75	15.0	20.00%	1.20	0.40	3.00
29	5	95	20	80	15.0	18.75%	1.19	0.25	4.75
30	0	100	15	85	15.0	17.65%	1.18	NA*	NA*
31			10	90	Not poss				
32			5	95	Not poss				
33			1	99	Not poss				

*Not applicable because formula places a zero in the denominator of the equation.

cession increases minority representation among drivers stopped and in the benchmark and thus decreases Caucasian representation among both categories.

For low levels of minority representation (the top of Table 12.8), the relative percentage difference can be very high—misleadingly high—even when the absolute percentage difference is low (in these cases, 2 percent). For Department 2, minorities comprise 3 percent of the drivers stopped and 1 percent of the benchmark population; the absolute percentage difference of 2 percent is paired with a relative percentage difference of 200 percent. Similarly, the disparity index for minorities and ratio of disparity are very high at 3.0 and 3.06, respectively.

In Table 12.9, like Table 12.8, we set the absolute percentage difference to a single value for all entries. In this table it is 15 percent—meaning the representation of minorities among those stopped is 15 percent higher than the representation of minorities in the benchmark. Note that for six departments, a 15 percent absolute disparity is impossible to achieve. For Department 3, for example, Caucasians comprise 90 percent of the stops, and minorities comprise 10 percent of the stops. Only the absurd benchmark results of 105 percent Caucasians and –5 percent minorities would produce a 15 percent absolute percentage difference. Similarly, an agency with 90 percent minority representation in the benchmark (Department 31) can never produce a 15 percent disparity.

Table 12.9 illustrates a problem that can arise for a researcher comparing levels of disparity across multiple departments or multiple areas. The choice of a cut-off point for identifying "problem areas" can lead to an inability to interpret the data for a jurisdiction. In Table 12.9 consider the cells labeled "not possible." Setting a cut-off using a 15 percent absolute percentage difference means that these departments are precluded (by their high or low percentage of minorities among stops or among the benchmark population) from being "eligible" for "problem area" status. It is impossible for those agencies to produce an absolute percentage difference of 15 percent.

Even when absolute percentage difference is kept constant in Table 12.9, the values of other disparity measures vary considerably. Departments with relatively few minorities in the stop and benchmark populations have relative percentage differences, minority disparity indexes, and ratios of disparity that are very large. Take, for instance, Department 5. Minorities represent 15.5 percent and 0.5 percent of the stop and benchmark populations respectively. Those two figures produce an absolute percentage difference of 15 and a huge relative percentage difference: 3000 percent. Department 6 has minority representation among drivers stopped and among the benchmark population that is a mere 0.5 percent greater than the minority representation in Department 5, but the relative percentage difference for Department 6 is half as small: 1500 percent. These very high relative percentage differences at the top of Table 12.9 are paired with similarly high minority disparity indexes and ratios of disparity. Departments 5 and 6 have very high minority disparity indexes of 31 and 16, respectively, and very high ratios of disparity of 36.50 and 18.86, respectively.

Low minority representation produces high disparity values; high minority representation produces low disparity values. In Table 12.9, the same absolute percentage difference of 15 percent produces the lowest disparity indexes in the departments with the highest minority representation. The lowest disparity index of 1.18 is for Department 30; its stops are composed of 100 percent minorities, and its benchmark population is composed of 85 percent minorities.

The relative percentage difference and the minority disparity index go from high levels at the top of Table 12.9 to low levels at the bottom of the table. Values for the ratio of disparity, however, fall and then climb, beginning at Department 23. Similarly, in Table 12.8 the ratio of disparity declined and then began to increase.

Measures of disparity are least stable and, correspondingly least likely to track each other, when the percentages of minorities in the stop and/or benchmark populations are very high or very low. Figure 12.2 illustrates this. It presents the results set forth in Table 12.8; absolute percentage differences were set to a constant—2 percent. The shape of the lines relative to each other—and not the distance between them—is the key to understanding the information conveyed in this figure. (Note that relative percentage differences that range between 200 percent and 2.04 percent in Table 12.8 are conveyed in the figure as ranging from 2.0 to 0.02.) Three of the measures—the relative percentage difference, disparity index and ratio of disparity—track closely, starting at the left edge until the ratio of disparity rises at the right side. The absolute percent difference—held constant at 2 percent—tracks all three of the other measures in the middle of the figure where they all manifest a relatively straight line. The absolute percentage difference continues to track the disparity index and relative percentage difference—but not the ratio of disparity—to the right side of the figure. Thus we can see from this figure an illustration of our main point: different measures of disparity can produce very different results for the same data.

USING CONTINGENCY TABLES TO IDENTIFY DISPARITY

Some researchers (for instance, Lamberth 2003b, Institute on Race and Poverty 2003, Engel 2004) have used contingency tables (or "crosstabulations") to assess the relationship, if any, between the race/ethnicity of drivers and various actions by police such as stops, searches, and dispositions. We provide general information here for the readers who are already familiar with contingency tables and the measures of association that can be used to interpret the findings.

Table 12.10 portrays search data from Tables 12.3 and 12.4 in contingency table format. Consistent with convention, the independent variable, the race/ethnicity of the driver, defines the columns, and the dependent variable, whether or not a search was conducted, defines the rows. Column percentages sum to 100 percent, and we read the table across. Searches were conducted of 17.61 percent of the stopped African Americans, 11.58 percent of the stopped Hispanics, and 7.00 percent of the stopped Caucasians.

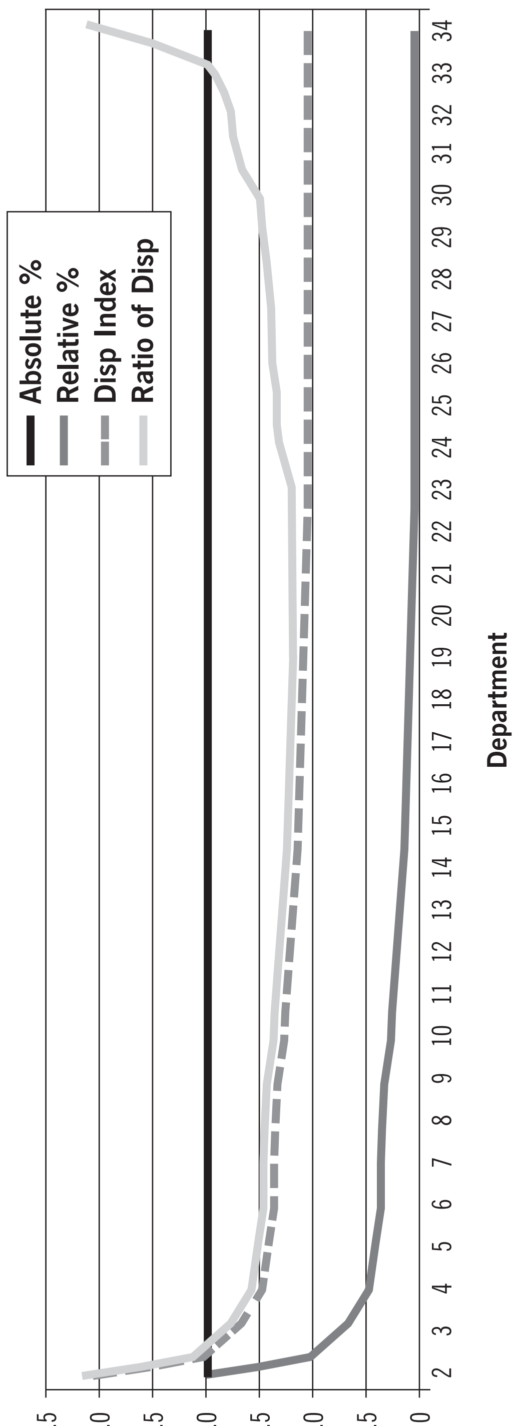

Figure 12.2. Four Disparity Measures for Multiple Departments When Absolute Percentage Difference is Set at 2

Source: Table 12.8

Table 12.10. Contingency Table to Assess Relationship Between Driver Race/Ethnicity and Police Searches

Driver Race/Ethnicity				
Search Activity	African Americans	Hispanics	Caucasians	Total
No Search	7,249 82.39%	5,919 88.42%	61,184 93.00%	74,352 91.48%
Search	1,549 17.61%	775 11.58%	4,605 7.00%	6,929 8.52%
Total	8,798 100.00%	6,694 100.00%	65,789 100.00%	81,281 100.00%

Note: The Contingency Coefficient is 0.121.

Researchers can use statistical programs (for instance, SPSS) to determine whether a relationship exists between the two variables and, depending on the technique used (for instance, Pearson's product-moment correlation), the direction and strength of that relationship. Because the two variables—race/ethnicity and whether or not a search was conducted—are both nominal, we produced a Contingency Coefficient, which is a measure of association based on chi-square. Its value of 0.121 indicates a weak association between the two variables. In other words, there is a weak association between a driver's race/ethnicity and whether or not the driver is searched.[8] Importantly, measures of association (and tests of statistical significance) provide information regarding disparity, not bias.[9]

8 The value of the Contingency Coefficient ranges between zero and 1, with zero indicating no association between the row and column variables and values close to 1 indicating a high degree of association between the variables.

9 As will be explained, tests of statistical significance used on vehicle stop data are useful as descriptive tools but do not allow researchers to generalize to a population.

For instance, if we had found a strong association indicating that African Americans were disproportionately represented among drivers searched, we would know only that a disparity exists, not why it exists. We cannot conclude that bias influenced search decisions because other factors could have caused the disparity.

Contingency analysis can be used when a benchmark can be conveyed in numbers, and not just in percentages. For example, contingency analysis can be used when the researcher can compare the number of people within each racial group stopped to the number of each group represented in the benchmark. Lamberth (2003b) conducted contingency analysis using the number of each racial/ethnic group observed by his stationary observers (the benchmark or denominator data) and the number of each racial/ethnic group stopped (the numerator data). The Institute on Race and Poverty (2003) used the number of minorities stopped and the number of minorities in the residential population as represented by the census. If a benchmarking method describes racial/ethnic representation in terms of percentages only (for instance, 20 percent of the benchmark population is minority, 80 percent is Caucasian), contingency analysis is not appropriate. Actual numbers of drivers in the benchmark populations, rather than percentages, is needed for contingency analysis.

USING MULTIVARIATE ANALYSIS TO IDENTIFY DISPARITY

In *bivariate analysis*, researchers look at the relationship between two variables. In the context of examining vehicle stop data, the researcher could look, for instance, at the relationship between the race/ethnicity of the driver and whether or not police conducted a search during the stop.

Multivariate analysis examines the impact of multiple factors (independent variables) on an outcome (the dependent variable). Multiple variables are taken into consideration, and the strength of the relationship between each independent variable and the dependent variable is determined by controlling for the impact of the other variables in the equation. Engel et al. (2004, 12) provide a description of multivariate methods as they apply to vehicle stop data:

> A multivariate statistical model is one that takes many different factors into account when attempting to explain a particular behavior. Unlike a bivariate model, it does not simply assess the relationship between two variables. Rather, a multivariate model examines many variables simultaneously, and therefore provides a more thorough and accurate interpretation of the data. For example, without controlling for the behavior of drivers, it is impossible to say whether higher rates of citations issued to particular drivers are justified based on legal considerations. A multivariate model can provide this information because it statistically controls for the existence of other variables in the model.

Smith et al. (2003) and Tomaskovic-Dewey, Wright, and Dzaja (2003) analyzed information from a survey of drivers in North Carolina, including information on the extent to which the drivers were stopped by police (see Chapter 10). These researchers wanted to find out whether the driver race/ethnicity affected the extent to which people were stopped. The frequency of being stopped during the reference period was the dependent variable. A bivariate analyses with these data would look at the relationship between the race/ethnicity of the survey respondents and the number of stops by police they reported. Researchers would not know from this bivariate analysis, however, if variables like driving quantity, quality, or location had affected stopping decisions by police. Researchers could show whether disparity existed (for instance, they might find that minorities were stopped more than Caucasians), but they would not know if race—or alternative, legitimate factors—produced that disparity. If a survey data set included information on driving quantity, quality, and location, researchers conducting multivariate analysis could look at the effect of race on the frequency of being stopped, controlling for those other factors.[10]

Clearly, multivariate statistical methods are superior to bivariate methods, but they can be used only on certain subsets and types of vehicle stop data, and they do not overcome the need for information on the alternative, legitimate factors that might influence stop and poststop activity by police.

The Types of Analyses that Can Be Conducted with Multivariate Methods

Researchers should not use multivariate methods to analyze the incident-level data collected on a police-citizen contact data form for the purpose of examining the relationship between race/ethnicity and stopping behavior by police.[11] If a researcher wants to examine whether race/ethnicity can be associated with the occurrence of X, the researcher needs data on those who experienced X and those who did not. The stop data collected by jurisdictions provides incident-level data only for the drivers who were stopped (see McMahon et al. 2002). This limitation of vehicle stop data does not apply to survey data (which includes information on people who were stopped and people who were not stopped). It also does not apply to poststop data (which includes, for example, information on those who experienced searches and those who did not). This is also true of other post stop decisions, including those related to stop disposition.

Incident-level data have been used by researchers to examine whether a search or a particular type of search (for instance, consent search) was conducted (for instance, Edwards et al. 2002a, 2002b; Schafer, Carter, and Katz-Bannister 2004; Withrow 2002; Lovrich et al. 2003; Smith and Petrocelli 2001); dispositions (for instance, Edwards et al. 2002a and 2002b; Schafer, Carter, and Katz-Bannister 2004; Cox et al. 2001; Crawford 2000; Engel et al.

---

10 The North Carolina team of Smith et al. (2003) was able to include some measures related to these constructs.

11 Below we'll contrast incident-level data with area-level data.

2004); length of stop (for instance, Withrow 2002); and whether the person was asked to exit the vehicle (for instance, Edwards et al. 2002a, 2002b). Independent variables in these equations have, of course, included the driver's race/ethnicity. Other independent variables have included the driver's age (Smith and Petrocelli 2001; Schafer, Carter, and Katz-Bannister 2004; Edwards et al. 2002a, 2002b; Lovrich et al. 2003; Engel et al. 2004), the driver's gender (Smith and Petrocelli 2001; Edwards et al. 2002a, 2002b; Engel et al. 2004; Schafer, Carter, and Katz-Bannister 2004), the reason for the stop (Schafer, Carter, and Katz-Bannister 2004; Engel et al. 2004), the geographic location of the stop (Lovrich et al. 2003), vehicle characteristics (Engel et al. 2004), roadway type (Engel et al. 2004), crime rate in the area of the stop (Smith and Petrocelli 2001), demographic makeup of the area of the stop (see Smith and Alpert 2003), number of violations detected (Lovrich et al. 2003; Engel et al. 2004), seriousness of violations detected (Lovrich et al. 2003), officer characteristics (Crawford 2000; Smith and Petrocelli 2001; Lovrich et al. 2003; Engel et al. 2004), and the day/time of the stop (Engel et al. 2004).

In the multivariate analyses described above, the unit of analysis used by the researchers was individual stop incidents, each of which was reflected in a single form completed by officers. Some researchers examining vehicle stop behavior have applied multivariate methods to a different unit of analysis—namely, to geographic areas. Using an area as the unit of analysis, researchers can conduct analysis of both stop and poststop data.[12] Area-level dependent variables used by researchers

---

[12] Researchers can analyze stop decisions at the area level using multivariate methods (something they can't do at the incident level) because with area-level data the researcher can estimate the population that was not stopped. With area-level data, the researcher has the number and racial breakdown of stops (X) and similar information (for instance, produced by adjusted census data) that describes the number and racial breakdown of people in the area at risk of being stopped. The people in the area composition who are not in the stopped population are the people who were not stopped; that is, they are the people who did not experience X.

include volume of stops (Smith et al. 2004; see also Smith and Alpert 2003; Petrocelli, Piquero, and Smith 2003), counts of stops of particular racial/ethnic groups (Zingraff, Smith, and Tomaskovic-Devey 2001), stops per 1,000 population 16 and over (Smith 2000), African American stops per 1,000 in African American population 16 and over (Smith 2000), and percentage of total stops that resulted in a search (Petrocelli, Piquero, and Smith 2003). Area-level independent variables used by researchers include crime rates (Spitzer 1999; Smith 2000; Petrocelli, Piquero, and Smith 2003); area demographics such as race, income, education (Spitzer 1999; Smith 2000; Petrocelli, Piquero, and Smith 2003; Engel et al 2004); traffic/travel patterns (Engel et al. 2004); area character such as whether it is characterized by retail or tourist business (Cox et al. 2001); proportion of drivers and proportion of drivers in accidents of a particular racial/ethnic group (Zingraff, Smith, and Tomaskovic-Devey 2001); and demand for police services (Parker 2003).

Below we describe three studies that incorporated multivariate methods. The first used incident-level data, the second used area-level data, and the third used both.

- Using logistic regression analysis and two years of data collected by an unnamed jurisdiction, Schafer et al. (forthcoming) examined the effect of driver characteristics (that is, race, gender, age) and stop characteristics (for instance, reason for the stop) on five stop outcomes: whether or not a search was conducted, whether or not a consent search was conducted, whether a discretionary versus nondiscretionary search was conducted, whether a search produced contraband, and whether the officer invoked a formal sanction versus providing only a warning.

- Smith et al. (2004) used regression to analyze the data collected by the Charlotte-Mecklenburg Police Department. Taking census block groups as its unit of analysis, the team developed models to predict the level of vehicle stops of minorities, the level of pedestrian stops of minorities, consent searches during pedestrian stops, and consent searches during vehicle stops. Independent variables for each geographic area included, but were not limited to, demographic composition, minority involvement in traffic accidents, calls for service in response to violent crimes, and calls for service in response to incivilities. Smith et al. 2004 also conducted multivariate analyses to assess whether the level of police activity in areas was justified by demands for service.

- Engel et al. (2004) used multivariate methods to analyze incident-level and area-level data collected by the Pennsylvania State Police. The team examined the impact of driver characteristics and stop characteristics on four outcomes (warnings, citations, searches, arrests). In its "hierarchical analysis," the team examined incident-level data within the context of the geographic area (municipality) of the stop. Independent variables at the municipal level included driving age population, percent male in driving-age population, percent African American in driving-age population, percent Hispanic in driving-age population, average commute (in minutes), and "three factor scores, measuring the latent variables poverty, residential mobility, and traffic/travel patterns" (p. 287). Independent variables at the incident level included driver characteristics (race/ethnicity, gender, age, residency), vehicle characteristics (for instance, registration, in/out of state, number of passengers), stop characteristics (for instance, time of day, day of week, roadway type), legal characteristics of the stop (for instance, reason for the stop, number of reasons for the stop), and trooper characteristics (for instance, gender, race, experience, assignment, rank).

The Key Limitation of Multivariate Analysis

Multivariate analysis is an important tool for social science and can have value for an examination of racial bias in policing. It does not, however, overcome the challenges associated with analyzing vehicle stop data—particularly those challenges associated with identifying and measuring the alternative legitimate factors that can influence police decision making. Multivariate analysis is based on certain assumptions, and a key one is "no specification error." This is a fancy phrase used by statisticians to reference a key theme of this book: for a method to be most effective it must take into consideration all of the alternative legitimate factors that might have an impact on police behavior. For multivariate analysis to be effective in determining whether driver race/ethnicity has a causal impact on police behavior, it must include independent variables that reflect the alternative legitimate factors that affect police behavior.

A researcher might find a significant relationship between independent variable X and dependent variable Y that would disappear if the researcher had included variable C in the model. A simple example illustrates this point. Let us imagine that a researcher finds a significant positive relationship between the consumption of high-grade coffee and the square footage of homes. Subjects who drink high-grade coffee, the researcher finds, are more likely to live in large houses. Clearly, drinking high-grade coffee does not cause a person to have a large house. The "omitted variable" C, which is wealth, leads to both the drinking of high-grade coffee and the purchase of large houses. Without the independent variable C in the model, the results are misleading: the results indicate a direct relationship where none exists. With wealth in the model, the multivariate methods would indicate a relationship between wealth (not high-grade coffee) and large houses.

Applied to vehicle stops, multivariate analysis can similarly identify a misleading relationship between the dependent variable and the independent variable. It is misleading because the inclusion of a previously omitted variable can make the relationship or correlation disappear. For example, multivariate analysis might find a relationship between race/ethnicity and police dispositions that would have disappeared (as it did in the analysis of the Washington State Patrol data) if the researcher had included number of violations or seriousness of offense(s) as independent variables. Not including key variables in a multivariate equation can also serve to "mask" racial bias. A researcher may find no indication of racial disparity in search decisions—where, in fact, it exists—because the researcher fails to include in the equation crucial independent variables. Knowles, Persico, and Todd (2001, 204–5) describe this danger (a specification error) in multivariate analysis of search decisions:

> If race has no explanatory power in the regression, this is taken as evidence of no discrimination (see, e.g., expert witness testimony by John Donohue in the case *Chavez v. Illinois State Police* [1999]). The drawback of this type of test for discrimination is that it requires data on the full set of characteristics that a police officer uses in deciding whether to search a motorist. If some characteristics were missing from the data, then race could have explanatory power due to omitted-variable bias. If race were found to be insignificant, there is still the possibility that police target individuals with certain characteristics because those characteristics are correlated with race and not because they are good predictors of criminality. Conditioning on those characteristics may lead to the wrong conclusion that race did not affect the search decision. Thus the validity of this type of test for discrimination hinges crucially on judgments about what constitutes a set of admissible conditioning variables and on whether the analyst has access to the full set of variables.

It is important to note that "specification error" is not unique to the analyses of vehicle stop data. Quite frequently, social scientists cannot identify or measure all of the factors that they should or would like to include as independent variables. This is not the noteworthy problem we are describing. The problem is irresponsible reporting of the results of multivariate analysis. The researcher should make explicit reference to the potentially relevant variables that were not included in the equation and report that these omissions could have had an impact on the results.

WHEN DOES DISPARITY MEAN BIAS?

In this chapter we have explained how to calculate various measures of disparity. Throughout this book we have discussed the challenges of isolating the causes of disparity. An identified "amount" of disparity in stopping behavior by police could be caused by any of the following: bias on the part of police; demographic variations in the quantity, quality, and location of driving; demographic variations in other legitimate factors that have an impact on police behavior; and/or other measurement error. The quandary for researchers is that they don't know what proportion of the disparity comes from what source. With strong benchmarking methods, researchers can *reduce* the number of plausible causes, but only in a perfect world where they can control for *all* alternative, legitimate factors and achieve perfect measurement could they equate a disparity measure or measures with police bias. For this reason, there is no agreed upon "bright line" researchers can set whereby disparity levels above it indicate racial bias and disparity levels below it indicate none.

Some researchers have set cut-off points (for example, Lamberth 2003b, 2004). These researchers are, in effect, arguing that if the disparity is particularly large, then, chances are, the alternative factors cannot explain all of it. Certainly, it is probably safe to say that the larger disparities are more likely than the smaller disparities to encompass many causes, including bias. It is important to note, however, another possibility: a large disparity could be produced entirely by alternative, legitimate factors, and a small disparity could be entirely produced by bias. Also, recall Myth 1 from Chapter 2: *No racial/ethnic disparity means no racially biased policing*. Indeed, the finding of no disparity does not prove lack of racial bias.

All disparity measures must be interpreted in light of the strength of the benchmark because it is reasonable to assume that the weaker the benchmark, the larger the potential influence of the nonbias factors. A disparity measure should not be interpreted the same way for strong and weak benchmarks. Lamberth and his team use an "odds ratio" (another name for what we call the "ratio of disparity") for measuring disparity based on the observation method of benchmarking. He writes, "We have taken the position that odds ratios between 1 and 1.5 are benign, and that odds ratios of 1.5 to 2.0 suggest that in the absence of other explanations, targeting of [minorities] may be occurring. Benchmarks of over 2 should be seriously considered by the [P.D.]" (Lamberth 2004, 25). One might accept Lamberth's chosen cut-off points for interpreting results from his observation benchmarking (a relatively strong benchmarking method) and still reject those same cut-offs for a study based on census benchmarking (a weaker method). All disparity measures must be interpreted in light of the strength of the benchmark.

Setting a Cut-Off Point

As noted earlier, there are no precise "rules of thumb" to help researchers answer this question: "At what level and under what circumstances does disparity equal bias?" Of course, executives of law enforcement agencies, other policy makers, and resident stakeholders are putting pressure on researchers to come up with "bright lines." They want an easy answer to the question: Is racially biased policing occurring in my jurisdiction or not?

Important for the researcher to understand is that setting a cut-off point is rather arbitrary. The researcher is guessing at the unknowable: How much of the disparity that has been detected between the racial/ethnic profile of drivers stopped by police and the racial/ethnic profile of the benchmark population is due to measurement error and unmeasured variables that influence police behavior? The arbitrariness of this enterprise is conveyed by McMahon et al. (2002). They compare the results of the Connecticut team of researchers (Cox et al. 2002) and the North Carolina team of researchers (Smith et al. 2003). Using examples, they show that if the Connecticut team had used the measures of disparity and cut-off points selected by the North Carolina team and the North Carolina team had used the Connecticut team's measures and cut-off points, the conclusions of each team would likely have been quite different. McMahon et al. then compare conclusions made by various researchers. Although researchers in Connecticut did not consider the difference between 12.3 (percent of African Americans in the stop population) and 9.1 (percent of African Americans in the benchmark population) to be meaningful in their results (Cox et al. 2001), the difference between 6.27 (percent Hispanics in the population over 18) and 6.71 (percent Hispanics among people stopped) was found to be meaningful by researchers evaluating St. Paul data (Institute on Race and Poverty 2001, 6).

For the researcher who chooses to select a cut-off point, we recommend the following, when it is feasible: (1) select the cut-off point before you analyze your results; (2) set the cut-off point in conjunction with a police-resident advisory board after educating that board about the challenges of drawing conclusions about police bias from calculations of measures of disparity; and (3) convey your results in a responsible manner to the public. The report should discuss the difficulty in interpreting the meaning of disparity. Lamberth's wording associated with the second level of his scale is constructive in this regard: "odds ratios of 1.5 to 2.0 suggest that, in the absence of other explanations, targeting of [minorities] may be occurring" (Lamberth 2004, 25). By referencing "other explanations," Lamberth is acknowledging that some explanations for police behavior were not accounted for with the methods he used. He is also inviting what we describe below as a qualitative review of quantitative data.

Choosing Not to Set a Cut-Off Point

A researcher might reasonably choose not to select a cut-off point, believing it unwise to select a point above which "a problem" is indicated or a "next step" is advocated. The Northeastern University team in its analysis of data for 361 Massachusetts agencies selected not to set a cut-off point that might indicate when the disparity was "too much." The following reason given by the researchers for their decision is a good one:

> [I]t is difficult to determine the appropriate threshold at which disparities become meaningful. Various standards have been used in other studies to draw conclusions about racial profiling based on comparisons between the demographics of those stopped and the demographics of those in the comparative population, but as a recent report by the Office of Community Oriented Policing Services (COPS) states "current research has failed to establish a consistent set of criteria to determine the nature and extent of racial profiling" [citing McMahon et al. 2003, 39]. As with other studies, we faced a problem of establishing a "bright line" above which the conclusion is that all departments are engaged in disparate citation practices that constitute racial profiling and below which all departments are not engaged in disparate citation practices. . . . In studies of disparity, regardless of topic area, it is generally inappropriate to conclude that any difference between the studied population and the comparative population automatically constitutes a meaningful disparity or racial bias. Such differences may be the result of real differences or may be a product of sampling or measurement error. Different studies rely on various thresholds above which they determine that observed differences are not solely attributable to error or chance. These thresholds differ dramatically depending on the type of sample used and the analytic methodology employed (Farrell et al. 2004, 15).

They conclude their discussion with this summary of their purpose and a reminder of the need for dialogue: "Understanding the limitations of establishing definitive measures of racial profiling, we instead seek to simply identify disparities… for each jurisdiction and identify those agencies that have the greatest levels of disparity when compared to other Massachusetts law enforcement agencies. . . . How much disparity is acceptable to a community is fundamentally a question that should be addressed by stakeholders and policy makers in each jurisdiction" (Farrell et al. 2004, 16).

Correctly, Cox et al. (2001, 16) acknowledge that "there are no measurable and objective specifications for determining what constitutes the practice of racial profiling by a police agency." Therefore, the stated purpose of their report, they say, "is to provide straightforward summaries of the traffic stops statistics." They explain that they "cannot arrive at an absolute conclusion of the existence or nonexistence of racial profiling." The statistics they present are given "in a variety of formats to provide the reader with sufficient information for identifying issues related to traffic stops."

When grappling with the question of "how much [disparity] is too much," researchers can avail themselves of two important tools. First, they can compare relative disparities (in the Northeastern University case, comparing disparities across jurisdictions). Second, they can encourage police and resident stakeholders to meet to discuss the data and what it means and does not mean. We will explain both of these tools after we comment on the use of tests of statistical significance.

Tests of Statistical Significance

Tests of statistical significance have limited application in studies of vehicle stops. These tests are usually used to make inferences about whether the results from a sample can be generalized to the population from which that sample was randomly drawn. The survey subjects of the North Carolina team were randomly sampled from the population of North Carolina drivers, and so it was appropriate for the team to discuss its results in terms of statistical significance. However, most data that are studied to assess the existence of racial bias represent information (gleaned from forms) on all police stops made in a jurisdiction, not a random sample. Because these data do not meet the underlying assumptions required for inferential statistical analysis, tests of statistical significance must be used with caution and primarily for descriptive analyses and not for purposes of generalizing to a population.

ASSESSING RELATIVE DISPARITY

Understandably, the law enforcement agencies and other organizations producing reports on vehicle stops and those reading them (policy makers, residents, and other stakeholders) will be frustrated by the lack of a clear message about whether racial bias is influencing stop and poststop activity by police. Although, as discussed above, we cannot provide cut-off points to indicate when disparity equals bias, we can offer researchers some tools to assist with interpreting their data. In this section we show how researchers have used various methods to identify individuals or areas with "the most" disparity. These identifications can, at least, provide some focus for the agency's further exploration including the agency's "qualitative assessment of quantitative results," as discussed further below.

Chapter 8 described benchmarking with data for matched officers or matched groups of officers. When applying this "internal benchmarking" method to St. Louis Police Department data, the team of Decker and Rojek (2002) used standardized scores (or "Z-scores") to analyze their officer-level data. One of the measures across which they compared "similarly situated" officers was the percentage of their stops that were of African Americans. They translated these percentages into standardized scores (see page 149) to assist in the interpretation of the results. Recall that standardized scores have an average of 0, and each increment of 1 represents one standard deviation. These scores allowed the researchers to identify which officers were the "outliers." The representation of African Americans among the drivers stopped by these outliers was much more or much less than the representation of African Americans among the drivers stopped by the similarly situated peer officers. An agency using internal benchmarking and standardized scores could identify the officers (or units of officers) with the "most disparity" and initiate the review described in Chapter 8 that will determine whether there are explanations other than bias for the disparity.

Identification of outliers using Z-scores could be applied to various populations and various variables. A researcher could use

Z-scores to compare areas of a jurisdiction, agencies within a state, and units within a department. A researcher could compare these entities across percent minorities stopped, percent minorities searched, and so forth. As we explained in Chapter 8, the strongest application would involve comparisons of similarly situated units. Units would be grouped together because they match across factors that seem reasonably related to variations in levels of the outcome being examined, such as minority stops.

As noted above, the Northeastern University team did not set a cut-off point indicating a level above which disparity indicated racial bias and below which disparity did not indicate racial bias. Instead it compared law enforcement agencies in Massachusetts to determine the ones with the most disparity. For each agency the team indicated for each of four measures of disparity whether any disparity was indicated. The four measures were absolute percentage differences for (1) citations of minority residents compared to the representation of minorities in the residential population, (2) citations of minorities compared to the representation of minorities in the estimated driving population, (3) the percent of minorities searched compared to the percent of nonminorities searched, and (4) the percent of minorities receiving citations (versus warnings) compared to the percent of nonminorities receiving citations. To indicate which agencies had the most disparity, the Northeastern University team (Farrell et al. 2004) calculated the median for the positive values for the two citation measures and reported which agencies had disparity levels above the medians for each measure.[13] Although these medians for the four measures could be considered "cut-off points," the Northeastern University team used these points only to describe where levels of dispar-

13 Some absolute percentage differences were negative, indicating that minorities were under-represented among, for instance, drivers stopped relative to their representation among the benchmark; for most agencies in Massachusetts, the absolute percentage differences were positive. The median was calculated based only on the positive absolute percentage differences.

ity were highest. It did not interpret above-median levels as a bright line indicating racial bias.

The Northeastern University method could be applied to other units for which the researcher has disparity measures—including officers, department units, and subareas. The researcher might choose absolute percentage differences, as did the Northeastern University team, or one or more other measures of disparity for purposes of the comparison. The researcher could rank the units based on these measures or select, as did the Northeastern University team, a *descriptive* cut-off point indicating which units have the highest levels of disparity.

The team conducting the analyses for the Charlotte-Mecklenburg Police Department in North Carolina (Smith et al. 2004) used advanced statistical techniques to identify jurisdiction block groups that had higher (and lower) than expected stops of minorities after *controlling* for key variables that might reasonably affect levels of minority stops.[14] As noted in the previous section of this chapter on multivariate analyses, the team predicted various dependent variables, including number of African American drivers stopped. "The important predictors for the number of African American drivers stopped were the number of white drivers stopped, the resident African American population, the number of African American drivers in accidents, as well as the number of successful consent searches in the vehicular context." Using Ordinary Least Squares regression, the team was able to explain 82.5 percent of the variance (adjusted R^2). This means that a large amount of the variation in police behavior across block groups was explained by legitimate (nonbias) factors.

Smith et al. (2004) used the results of the regression analysis—specifically, the coefficients for each independent variable—to determine for each block group how many African

[14] This team—Smith, Davison, Zingraff, Rice, and Bissler (2004)—is the same team that conducted the analyses of data for the North Carolina State Highway Patrol referenced throughout this book.

Americans should have been stopped. They then compared the "predicted number of African American drivers stopped" to the "actual (or observed) number of African American drivers stopped" for each block group. Their results are plotted in Figure 12.3. The center line is the regression line where a block group would fall if its predicted and observed stops matched perfectly. The line above and the line below the center line represent the 95 percent confidence interval. The boxes above the regression line represent "areas with higher than expected numbers of African American drivers stopped"; the ones above the top confidence interval line were defined as "outliers." Boxes below the regression line represent areas with "fewer African Americans stopped than expected." Boxes below the lower confidence interval line are also "outliers."

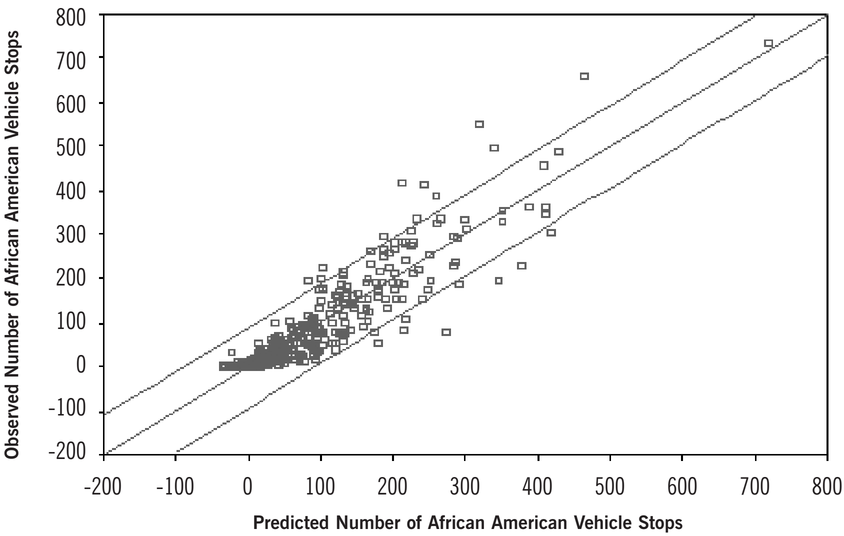

Figure 12.3. Results Used to Identify Charlotte-Mecklenburg Block Groups with Higher or Lower than Expected Numbers of African Americans Stopped

Source: Smith et al. (2004, 86).

Responsibly, Smith et al. (2004) admit that they cannot know whether or not the areas above the regression line in the figure are areas with "excessive" stops of African Americans. They say in their report only that there are more than expected stops of African Americans for unknown reasons. Even though the team did not draw conclusions about the existence or lack of racial bias, its analyses provided sufficient direction for discussions of the data. In an attempt to understand what factors might account for the positive and negative outliers, the team discussed possible explanations for "outlier status" with the leadership in the relevant police districts. In the second step, the team reports, the "citizens' advisory committee, along with representatives of the CMPD, will discuss our findings, and make any necessary decisions about whether—and what—corrective measures are needed" (Smith et al. 2004, 25).

The examples provided above represent methods for assessing relative disparity. They allow the researcher to identify officers, units or areas that have the most disparity. Such an identification can serve to help policy makers identify the high priority targets for additional review or for change efforts as discussed more fully in the next chapter.

QUALITATIVE REVIEW OF QUANTITATIVE DATA

By discussing its data with district commanders and by referring the results of its analysis to a police-community group for further discussion, the Charlotte-Mecklenburg team is conducting and promoting qualitative reviews of quantitative data. These reviews can help ensure that jurisdiction data are correctly and responsibly interpreted. Like in the Charlotte-Mecklenburg model, two reviews are advisable: (1) a review and discussion of the results by researchers and law enforcement agencies, and (2) a review and discussion of the results by law enforcement personnel and resident stakeholders.

Reviewing the Results with Law Enforcement Personnel

The independent researcher or researcher employed by the law enforcement agency should discuss the results of vehicle stop data analysis with sworn personnel before publishing them. The purpose of this discussion is to gather information from a "street perspective" regarding what the data mean. The purpose is not to "explain away" any disparity that may have been identified but to better understand what factors—legitimate or otherwise—might be producing the results.

The teams headed by John Lamberth meet with law enforcement personnel in the jurisdiction after the empirical results are produced and before the report is written. The results are shared, and the personnel in attendance are asked to discuss why some activities (stops, searches) or geographic areas (particular intersections) might indicate racial disparities. In one city the chief gave researchers a valuable explanation for the particularly high volume of stops at an intersection: a specific directive from the command level was issued to increase traffic enforcement at the intersection due to community complaints of speeding. In another city, one with a high rate of consent searches of African Americans, researchers learned from police that directed patrols conduct more consent searches and that they are assigned to high-crime, high-minority areas where citizens have requested "quality of life" enforcement. These are two examples of the value of reviewing results with police.

In discussions with command staff of districts concerning higher-than-expected vehicle stops of African Americans, the team analyzing the Charlotte-Mecklenburg data was given the following possible explanations for the positive outliers:

> 1) checkpoint activity (set up of a vehicle check point); 2) a rash of accidents in an area resulted in more patrolling; 3) presence of major north-south and of east-west thoroughfares; 4) proximity to the coliseum; 5) presence of a police substation; and 6) 'crackdown' area where drivers are 'stopped for everything' because of erratic driving (Smith et al. 2004, 87).

The team of Farrell et al. (2004) could not meet face-to-face with personnel from each of the over 350 Massachusetts law enforcement agencies involved in its study. Instead the team released a preliminary report for review. The team invited comments from law enforcement officials and community members, and it "held six regional community meetings" at which they "actively solicited . . . reactions to the findings" (Executive Summary, 2). Decker, Rosenfeld, and Rojek, who have been analyzing the data for the state of Missouri, similarly invited feedback. They sent preliminary report copies to the over 650 law enforcement agencies in the state.[15] Agencies were asked to check for errors, and they were offered a space in the narrative to provide their comments on the results.

As part of their study of vehicle stops in North Carolina, Smith et al. (2003) held focus groups with officers (as well as with citizens). Officers in the focus groups told researchers the factors that they consider in deciding to stop, to search, or to select a particular stop disposition. Their responses were used to add perspective to the empirical data.

Lawson and Fitzroy (2004) had a policy of open communication with officers and their supervisors throughout the entire study period of their research, including the data analysis phase. Early analysis produced bivariate results for each agency participating in this county-wide study. Participating chiefs and their staff were invited to review the crosstabulations and to share their insights (for instance, why the rate of stops was highest during certain time periods, why certain shifts were more likely to give more serious dispositions). The participants also suggested different ways to categorize the data (for instance, changing time periods to match with shifts).

---

15 The Missouri reports for 2000 through 2003 are available on the Web at http://www.ago.state.mo.us/racialprofiling/racialprofiling.htm.

Police-Resident Discussions of the Results

The Northeastern University team, in both its Massachusetts (Farrell et al. 2004) and Rhode Island (Farrell et al. 2003) reports, indicates that the ultimate interpretation of the results comes during discussions between police and citizens. As indicated above, the Charlotte-Mecklenburg team promoted such discussions as well. One benefit of including residents in discussions of results is the fresh and helpful perspective they bring to understanding what the data mean. Like the police, residents have information about the jurisdiction that can add perspective and context to the numbers produced by the researcher. But discussions between police and residents are about more than how to interpret data. The issue of racially biased policing has, in many communities, exacerbated the "divide" between police and residents, particularly residents who are racial/ethnic minorities. Data collection has the potential to help heal the divide and provide direction for joint reform efforts by police and community members. Police-resident discussions of data become a part of the change process and can become the vehicle for additional reform efforts. We discuss in the next chapter how police and residents can come together to use these data for purposes of reform.

CONCLUSION

In this chapter we discussed four ways to present the results of vehicle stop analyses and explained why conclusions about bias cannot be drawn from calculations of measures of disparity. With regard to the choice of a measure or measures to convey results, we argue that that are no "correct" ways of proceeding, but rather various options for researchers to consider, options with positive and negative aspects.

Disparity can be conveyed through absolute percentage differences, relative percentage differences, disparity indexes, and/or ratios of disparity. These measures can be used to present results on stops, searches, dispositions, and other types of vehicle stop data. The challenge is not in producing these measures of disparity but in deciding which one or ones to use and present.

Some social scientists use just one measure of disparity in their reports to reduce ambiguity and avoid multiple interpretations of results. Others prefer to report multiple measures of disparity. We explained both points of view in the chapter and emphasized that the conclusions drawn from one measure might be very different from those drawn from another—a possibility that must be explained in the report of findings.

Social scientists also disagree on whether it is advisable to select a cut-off point above which disparity levels are said to indicate racial bias. Researchers who advocate cut-off points argue that if the disparity is particularly large, then, chances are, the alternative, legitimate factors affecting stop and post-stop activity by police cannot explain all of it. These researchers argue that conclusions about the existence or absence of racial bias can and should be drawn from disparity measure calculations in order to provide the clarity needed to guide jurisdiction policy and practice.

Other researchers claim that any cut-off point is arbitrary—providing a false sense of clarity where none exists. They note that even large amounts of disparity could be wholly explained by nonbias factors. Those who do not favor the selection of cut-off points advocate that researchers provide the public with information that describes disparity. Descriptive information on disparities could be supplemented with comparisons of relative levels of disparity across various units under study (for instance, areas of the city, units of a department, jurisdictions within a state) and by the qualitative analysis of quantitative data. Conclusions about racial/ethnic bias as the cause of disparity must be evaluated in light of the strength of the benchmark and the extent to which nonbias factors have been addressed.

In the next chapter we expand on the coverage of the qualitative analysis of quantitative data by describing how police and resident stakeholders can come together, reflect upon the vehicle stop data analyzed by social science researchers, and identify methods for improving policing practices and the relationships police have with local residents.

XII Using the Results for Reform

Vehicle stop data have potential and constraints as a means of measuring whether policing in a jurisdiction is racially biased. Previous chapters have explained that the limits of social science preclude researchers from drawing strong conclusions regarding the existence or lack of racial bias. Faced with this fact, one well might ask: of what value are these results if researchers cannot report, with confidence, the existence or lack of racial bias in the jurisdiction? The answer is that they can be of significant value. These results can serve as a basis for constructive dialogue between police and residents, which can lead to (1) increased trust and cooperation and (2) action plans for reform. In its report on traffic stop data for the state of Rhode Island, the Northeastern University team wrote: “We do not view this analysis as an end of the discussion about the existence and extent of racial profiling in Rhode Island, but rather it will provide . . . information to begin an important dialogue. . . . [A] well conceived and implemented study of racial disparities in traffic stops can serve as a very useful springboard for community level conversations about the issues of racial profiling” (Farrell et al. 2003, 6).

Below we describe various ways that police and resident stakeholders[1] can come together to reflect on the results of data collection efforts. Their ultimate aim is mutual understanding and reform. Specifically, we describe in this chapter

- who should be brought together;
- what information—including vehicle stop and poststop results—this group might explore; and
- the types of changes the group might recommend.

As articulated by Chief John Timoney (2004) of the Miami Police Department, the reality is that "race is a factor in policing." Every police executive needs to consider and address the issues of racially biased policing and the perceptions of its practice. Because all agencies can make progress on this issue and because the data will never "prove" or "disprove" racially biased policing, we contend that vehicle stop data collection and analysis should never be viewed—either by police or resident stakeholders—as a "pass-fail test" (Farrell 2004). Instead, it should be viewed as a diagnostic tool to help pinpoint the decisions, geographic areas, and procedures that should get priority attention when the agency, in concert with concerned residents, identifies its next steps for addressing the problem or perception of racial profiling.[2]

---

1 In this chapter the term "resident stakeholders" refers to citizens, journalists, advocacy group members, government officials, and others who reside in the community and have a particular stake in the outcome of researchers' race data analysis.

2 This should not be construed as an endorsement of mandatory data collection. As indicated in the first PERF publication, there are pros and cons to data collection.

THE TASK FORCE AND ITS MEMBERSHIP

In Chapter 3, "Getting Started," we recommended that jurisdictions create a local racial profiling task force to guide police departments in the development of their data collection system.[3] This task force, composed of fifteen to twenty-five people, could plan how data would be collected and analyzed. The task force would bring credibility to the data collection system, and its members would understand both the limits and the potential of vehicle stop data analysis. We recommend including people in the community who are most concerned about racial bias and police personnel representing all departmental levels, particularly patrol.

It is preferable, but not essential, that the task force be convened before the data collection initiative begins. If it is formulated after data collection has started, however, it still has an important mission—engaging in constructive dialogue to identify targets for change efforts. It is important, however, that this group meet and begin its work *before* the report of findings on the vehicle stop data analysis is publicly released.

A group with equal representation of law enforcement personnel and resident stakeholders should review and discuss the data. Nonresident stakeholders also could be included. They could be representatives from state or national groups, such as the American Civil Liberties Union (ACLU), the National Association for the Advancement of Colored People (NAACP), and the Urban League; nonresident commuters to the jurisdiction; and nonresident owners of businesses located in the jurisdiction.

It is usually appropriate for the agency executive to call for and develop this task force. It then serves in an advisory capacity to the executive and makes recommendations that he or she will consider adopting. The agency executive should not be a member of the group since it has been convened to provide him

---

3 Because data collection was organized at the state level, the Northeastern University team had a state-level task force advising it. The team, however, advocates that discussions of the data occur at the local level.

or her with advice on what actions to take. We recommend, however, that the executive attend the task force meetings. By attending the meetings, the executive can convey to task force members, the executive's staff, and the wider community the importance of the issue. There may be circumstances when another official or group develops the task force rather than the law enforcement agency executive. For instance, a mayor or city council might call for a task force for a jurisdiction or a governor might convene a statewide task force. The executive should be a member of the task force if it was not set up and overseen by the executive; the members would make recommendations to the person or organization that developed their group.

The local racial profiling task force should meet on an ongoing basis. For some of the early discussions described below (for instance, on trust-building and on general issues and concerns related to racially biased policing), we advise the use of a trained, neutral (nonpolice, nonstakeholder) facilitator. This facilitator should have experience working with groups on issues that provoke emotions and passions and have knowledge of the topic of racially biased policing. This facilitator might be retained to oversee the long-term work of the task force or, after the early sessions, turn over meeting facilitation to a task force chair or to co-chairs. For the co-chair model, the group may elect, or have appointed, one co-chair who is an internal stakeholder and another who is an external stakeholder. This group may have a finite tenure or may become a permanent fixture in the jurisdiction.[4]

---

[4] For various reasons, a jurisdiction may be unable (or, unwilling) to convene a task force of police and stakeholders. In such circumstances, the department should convene personnel to discuss key topics outlined below, including general issues related to racially biased policing, the vehicle stop results, other sources of information, and needed reforms.

THE AGENDA OF THE POLICE-STAKEHOLDER TASK FORCE

The first few sessions of the task force (sessions led by a neutral facilitator, as explained above) should be devoted to developing trust between police and resident members. The task force then would discuss

- general issues and concerns related to racially biased policing,
- what the vehicle stop data indicate,
- what other sources of information indicate about racial bias and perceptions of racial bias, and
- reforms that could be implemented.

Developing Trust

In a rare situation, a stakeholder group may be able to begin its discussions of racially biased policing at the first meeting; most groups, however, will be well served by engaging in some exercises and discussions on topics other than racial bias before delving into the volatile topic that brings them together. A group in Lowell, Massachusetts—not a task force but a group formed for a one-time discussion—began immediately talking about racial bias. After some finger-pointing, raised voices, accusations by citizens against police, and defensiveness on the part of police, the group turned its attention to developing ways to resolve the particular problems it had identified. On their own, without prompting from the facilitator, the group members agreed that they needed to meet regularly to continue the process of sharing, listening, and resolving problems. Ed Davis, chief of the Lowell Police Department, continued the group as the "Race Relations Council," which the mayor later described as "the best thing that has happened in Lowell in a long time."

Although this particular group was able to move during a single session from the heated and angry exchanges at the beginning of the meeting on the controversial issue of race to a sober and rational discussion of a constructive plan of action, most groups cannot. We recommend that task forces engage in activities that will develop trust among members before tackling the challenging topics that define their existence. This trust-building may require a number of meetings.

The Chicago Forums

One trust-building model comes from Chicago. The former superintendent of the Chicago Police Department, Terry Hillard, sponsored a series of forums for police and minority residents of the community. Community activists were recruited to aid the police department in its search for solutions to racial tensions. Department staff of all ranks were also invited to participate. Before the first forum was convened, participants were surveyed for their opinions about racially biased policing and the department's strengths and weaknesses regarding minority outreach. In the survey, respondents also were asked for their ideas on how to improve relations between police and minorities and for their thoughts on how to resolve issues. A facilitator moderated the initial sessions.

During the morning session of the first forum, community members were asked to talk about strengths and weaknesses of their interactions with the police, and police staff were asked to listen and hold their responses until later in the day. Lunch was structured as a mixer, with informal discussions. In the afternoon, police staff shared their thoughts and reactions to the morning session, and residents were instructed to listen and not respond. Then there was an opportunity for discussion. While the issue of police racial bias was raised by both groups during this first meeting, it was just one of many issues raised. Race issues became a more central focus in subsequent forums and, during those gatherings, the group identified specific actions to be taken by both police and community members to address them. Superintendent Phil Cline who succeeded Hillard has continued these forums.

The Lamberth Workshops

John Lamberth's consulting team uses a two-session workshop to "enhance the trust between law enforcement and the local community" and to develop "collaborative community-based racial profiling solutions" (Clayton 2004). For the first gathering, the Lamberth team holds separate sessions with the police and resident stakeholder participants. The purpose of these separate discussions is to "enhance participants' understanding of the issues" surrounding racially biased policing.

Discussions within these separate groups address the definition of racial profiling, differing perceptions of the issue on the part of law enforcement and resident stakeholders, and the expectations and responsibilities of police and drivers during vehicle stops. By the end of the first session, each group has identified

- safety issues that concern police,
- concerns or fears that drivers might have when stopped by police,
- ways racial profiling harms police-community relations, and
- its expectations when making contact with the other group.

During the second session of the workshop, the group composed of police and the group composed of resident stakeholders are brought together to engage in small- and large-group discussions and activities. The police group and the resident stakeholder group review their separate discussions from session one and identify the areas where their expectations and perceptions are shared and where they are different. Together, the first session and first half of the second session serve to initiate constructive dialogue, develop trust between participating police and resident stakeholders, and identify common concerns and expectations. These sessions set the stage for the rest of the workshop during which the participants develop a plan of action for addressing issues related to racially biased policing and the perceptions of racially biased policing.

General Issues and Concerns Related to Racially Biased Policing

We have discussed the first item on the agenda of a local racial profiling task force: developing trust. We have described two examples of efforts to develop trust and enhance communication during non-stress times: forums convened in Chicago for police and minority residents and two-session workshops developed by John Lamberth's consulting team. Following trust-building gatherings similar to those we have described, members of the task force should turn to a general discussion of concerns and perceptions related to racially biased policing. To provide structure to the potentially heated conversation, the facilitator might invite resident stakeholders to share their concerns—allowing them to voice their perspective without defensive responses by the police. While the police might feel inclined to "explain away" all the concerns voiced by citizens (and, indeed, there will be incidents described by residents where the police feel strongly—and maybe correctly—that there is a race-neutral explanation), it will ultimately be more valuable for the police to just listen to the residents' concerns. Residents need to be heard on this issue and taken seriously. This discussion also can highlight for police how important it is to deal with *perceptions* that police in the jurisdiction are racially biased. Then the facilitator could ask police on the task force to share their concerns related to accusations or perceptions that bias is influencing their policing decisions.[5]

A Review of the Vehicle Stop Results

After a general airing of concerns, the task force should be ready to conduct a qualitative (that is, nonempirical) review of the

---

5 The task force should include police leaders at all ranks who are open to exploring the issue of police racial bias and committed to identifying ways of doing business that can reduce or prevent the problem and perceptions of the problem. These people should be problem solvers and consensus builders.

quantitative data on vehicle stop data (see Chapter 12). This review is a continuation of the data analysis. During the earlier empirical examination of the stop data, the researcher will not have been able to consider all of the factors that might have influenced stopping decisions by police. A "qualitative" review allows for a constructive assessment of the factors, other than bias, that might account in whole or in part for findings of disparity (or lack thereof). The police and residents who have been brought together on the task force have valuable knowledge about the activities of police, about residents, and about geographic areas in the jurisdiction. Therefore, they can provide a unique and helpful perspective for understanding the empirical results obtained by the researchers.

The goal of the qualitative review of quantitative data is not to determine whether the agency "passed" or "failed" a racial profiling test. As stated earlier, the goal is to identify geographic areas, procedures, and decisions that should get the highest priority when the police department initiates efforts to address community concerns. Even though the quantitative data cannot provide the whole picture or a perfect picture, the data, if carefully interpreted, can direct the task force toward particular reform targets such as stops of minorities for equipment violations, consent searches of young African American males, or vehicle stops on the "south side" of the city.

Before reviewing the data, members of the task force should become informed about what can and cannot be understood from the analysis of vehicle stop data. They can read *Understanding Race Data from Vehicle Stops: A Stakeholders' Guide* (Fridell forthcoming, 2005), a book that is a companion to this volume, or be otherwise educated (perhaps by the researcher) about key concepts, such as the meaning of "benchmarking" and the meaning of "disparity." Once all members of the group have a good preliminary understanding of vehicle stop analysis, they can review the stop and poststop data. The following questions can help guide this discussion of the data:

- Are there indications of disparity in the stop or poststop results?
- Are there reasons, other than racial bias, that might have led to these disparities? For what activities (for example, stops, searches, choice of disposition) is racial bias a possible or probable cause?
- Regardless of whether or not bias is a cause, what is the impact of particular disparities on residents and on relations between police and residents of the jurisdiction? Do the costs outweigh the law enforcement benefits?

Each of these questions will now be examined in greater detail.

An appropriate first question to guide the discussion of the vehicle stop data is "Are there indications of disparity?" It is important for the group to keep in mind that the discussion at this point is about indications of "disparity" not "bias."[6] This conversation about disparity may be shorter than later conversations about the other questions listed. The key is to summarize what disparities were identified by the empirical analyses.

The more interesting, challenging, and longer discussion will focus on the reasons, other than racial bias, that might have led to these disparities. This conversation might start by focusing on each specific finding of disparity (for instance, disparity in stops across racial groups in Area A). Participants might reflect on how the methods used to produce the measure did or did not capture certain important factors. For instance, a resident participant might point out that racial disparity in stops around a stadium that was identified using census benchmarking might reflect the high volume of nonresident, multiracial/ethnic traffic on game days. An officer might report that

---

[6] The group should also be reminded that the same methodological challenges that keep researchers from equating disparity with bias can produce results showing no disparity when racial bias does, in fact, exist (see Myth 1 in Chapter 2).

the high level of stops in a particular minority area is the result, at least in part, of requests from residents in that area for strong enforcement of the speed limit.

The purpose of the discussion of these on-the-ground realities is not to "explain away" disparities but to examine legitimate factors that might account, at least in part, for them. The task force will also consider the possibility that certain identified disparities could be the result of biased decisions on the part of police. If the group cannot identify alternative, legitimate explanations for findings of disparity, if there is an accumulation of disparity findings or very large levels of disparity, or if a particular police activity is highly discretionary and thus vulnerable to bias, the group should consider this possibility. The results—whether quantitative or qualitative or both—will never be definitive: this key point has been repeated often in the preceding chapters. However, the conversation can and should proceed despite these inevitable constraints. The group is not looking for "proof" of racial bias (if it is, it will not find it); instead it is trying to identify priorities for its initial change efforts.

Even if no causal linkage to racial bias is perceived by task force members, their deliberations may reveal the need for some changes in police procedures. Disparities produced by law enforcement activities may not necessarily implicate racial/ethnic bias (or, as is inevitable, the link is unclear), but the activities may be detrimental nonetheless. It is constructive for the group to discuss the potential negative impact on the jurisdiction of even these (potentially) race-neutral disparities and target efforts to change them. For instance, data on poststop activity by police may indicate that African Americans are much more likely than Caucasians to be asked to consent to a search; the data also may show that these consent searches are very unproductive (as measured by hit rates). Although it may not be possible to determine whether bias produced this disparity (see Chapter 11), the group may decide to recommend some changes nonetheless. Such a recommendation may make sense if minorities in the community perceive racial bias in these requests by police. This disparity in searches—regardless of whether it is caused by bias—may be too costly in terms of relations between police and minorities. The frustration and anger of minorities may be too high a price to pay for whatever crime control value is derived.

A Review of Other Sources of Information Regarding Racial Bias and Perceptions of Racial Bias

Task forces should consider, in addition to vehicle stop data, other sources of information when trying to identify positive steps that can be taken in the jurisdiction to address racially biased policing and perceptions of its practice. These alternative sources could include conventional wisdom regarding the types of law enforcement activities that might be most vulnerable to officer biases, surveys of jurisdiction residents to assess perceptions of racially biased policing, and results of focus groups held around the jurisdiction.[7] The group might want to review various other sources of data within the department (for example, aggregate data on official complaints against officers, data on the use of force, and arrest data).[8] Selected tapes from in-car video cameras might be another valuable source of information.

Proceeding to Reform without the Confession

The discussions outlined above can strengthen the police-community relationship and promote trust, as well as highlight areas of concern to guide reform efforts. These benefits, however, can be lost if the move from discussing results to discussing

---

7 In some jurisdictions, focus groups of residents might be supplemented by focus groups of nonresidents who nonetheless have a stake in the professional performance of police. These nonresidents might include business owners and commuters.

8 The department researcher within the Las Vegas Metropolitan Police Department examined force reports. In one analysis, he looked at the race and ethnicity of subjects who were cuffed during a stop and then released with no arrest.

reform is predicated on a forced "confession of guilt" on the part of the law enforcement department.

Following a discussion of the vehicle stop data by the task force, resident stakeholders in the group (including government leaders) may demand a confession of guilt. This is a mistake. A confession of guilt should not be a criterion for moving the discussions forward because vehicle stop data collection/analysis is not a pass-fail test. As conveyed throughout this book, a jurisdiction will not have "proof" of racial bias (or the lack thereof). Moreover, "proof" of racial bias is not a prerequisite for decisions that reforms are worthwhile. All agencies can move closer to the ideal of bias-free policing. Perhaps most importantly, exploring reform without a forced confession of guilt is the most constructive and effective way to proceed.

Police-stakeholder discussions of "racial profiling" that involve finger-pointing by residents and defensiveness by police are not helpful. Discussions when resident stakeholders accuse police of "widespread racism" and of frequently "stopping people solely on the basis of race" are not constructive. These types of accusations inevitably lead to defensive responses on the part of police. On the other hand, discussions along these lines could be constructive: stakeholders can acknowledge how racial/ethnic bias still is pervasive in our society and how even well-meaning people (including, but not limited to, police officers) might make decisions that manifest bias. The police in these communities can acknowledge the concerns of the community and express a willingness to engage concerned citizens in discussions about how to move forward.

Without making a confession, chiefs can still acknowledge the need to address the concerns of their constituencies. Chiefs might say that, while they cannot prove whether or not their agencies have a problem with racially biased policing, they do know that some residents have very real concerns and perceptions of a problem that must be taken seriously. The chiefs could acknowledge that these concerns and perceptions harm the relationship between the police and the racial/ethnic minorities in the community and could welcome a dialogue that leads to positive changes.

No agency executive should declare his or her agency "innocent of" or "immune from" racial bias. The many caveats in this book regarding vehicle stop data make clear why such a declaration is unwise. The results of vehicle stop data analysis will never support such a strong statement and, besides, it's very unlikely that any agency is without room for improvement on this issue. A statement of innocence would anger constituencies that have strong concerns and perceptions of police bias, and it can cause significant harm to police-minority relations. Furthermore, this chief can never implement reform measures with any degree of acceptance from agency personnel since he or she has previously declared publicly that there is no problem to address.

CHARTING CHANGE INITIATIVES

Having agreed to move forward without a public declaration of guilt or innocence by the law enforcement agency, the local racial profiling task force can begin outlining specific change initiatives, using as a guide its discussions of general concerns regarding racial bias, vehicle stop data that may indicate bias and/or deleterious disparity, and other sources of information. The interventions the task force identifies might be specific to a particular "finding," or they might be of a general nature. As an example of specific findings that can lead to reforms, the group may find in the data a large number of consent searches of minorities that are unproductive or a curiously large proportion of minority stops with unproductive searches and "no action" dispositions. To address the specific problem of many consent searches of minorities that are unproductive, the task force might suggest that the chief adopt an agency policy requiring that citizens sign a consent form before being searched. This consent form would inform residents of their right to refuse. Alternatively or additionally, the group might suggest that the chief implement a minimum "level of proof" for consent searches, such as reasonable suspicion.[9] In response to the finding of a large proportion of minority stops with unproductive searches and "no action" dispositions, the group might suggest that the agency executive revise or "retrain" on policies to ensure that stops are made only for legitimate reasons and establish means for commending officers whose searches are the most productive (as measured by their hit rates). Hit rates should not be examined in isolation, but rather within the context of other performance or productivity measures. To reduce questionable stops, the group might suggest that the agency adopt a policy that prohibits pretext stops.

These are a few specific changes that could be recommended. Broader initiatives are outlined in *Racially Biased Policing: A Principled Response* (Fridell et al. 2001). In that book the authors argue that all agencies—whether they have collected vehicle stop data or not—should consider efforts in the following areas:

- Supervision/accountability,
- Policies,
- Recruitment and hiring,
- Education and training, and
- Minority community outreach.

Community members should be full partners in implementing the solutions. For instance, residents could help develop the agency's policy on antibiased policing, assist with efforts to recruit minority officers, participate in the development of a recruit or in-

9 The reforms in this example were implemented by Chief Stanley Knee in Austin after vehicle stop data showed that greater proportions of minorities than Caucasians were subject to consent searches. The consent searches of minorities were not very productive, and resident stakeholders perceived that racial bias was the cause of this identified disparity. The chief implemented a consent form and a policy requiring reasonable suspicion on the part of the officer prior to requesting consent to search. He set a goal of increasing the level of productivity of consent searches.

service training curriculum, support agency outreach efforts to diverse communities, or identify external funds that could go toward the purchase of equipment and software to support efforts.

Specific findings or general conclusions regarding the problem might prompt the task force to recommend the collection of more information by the jurisdiction. For example, an agency that conducted analyses of the jurisdiction as a whole might choose to conduct subarea analyses to determine whether there are particular geographic areas where disparities are very high. An agency that used a relatively weak benchmark and found areas with large disparities might implement a stronger benchmark in the identified areas. An agency that compared its vehicle stop data to an "external benchmark" (for instance, agencies using observation benchmarking, benchmarking with adjusted census data, or benchmarking with blind versus not-blind enforcement mechanisms) might choose to implement internal benchmarking (see Chapter 8). The agency then could identify the particular officers who produced the disparity so that their policing decisions could be subject to further review. Alternatively or additionally, an agency might decide that additional data elements need to be included on its forms so that it could further explore a potential problem area (for instance, consent searches).

To better understand some aspect of the data, an agency may choose to conduct focus groups of officers, a community survey of perceptions of racially biased policing, or a consumer survey (for instance, a survey of drivers stopped by police). All of these initiatives would help the agency to obtain positive and negative feedback regarding community members' interactions with officers. However, the police and resident stakeholders on the task force should not emphasize data collection and measurement to the extent that they postpone or neglect the most important work—implementing change.

CONCLUSION

This book has set forth both the benefits and the limits associated with the use of vehicle stop data to measure whether policing in a jurisdiction is racially biased. "Benchmarking" is the method of analysis used to make this measurement, and, as noted in Chapter 2, benchmarking presents a real challenge for researchers because they must consider the following four alternatives to the bias hypothesis when analyzing data on drivers stopped by police:

- *Racial/ethnic groups are not equally represented as residents in the jurisdiction.*
- *Racial/ethnic groups are not equally represented as drivers on jurisdiction roads.*
- *Racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior.*
- *Racial/ethnic groups are not equally represented as drivers on roads where stopping activity by police is high.*

Researchers must similarly consider alternatives to the bias hypotheses when analyzing search, disposition, and other post-stop data. Identifying and ruling out the "alternative, legitimate factors" that can influence police decisions concerning stops, searches, or dispositions is a complex and painstaking task. Nevertheless, many departments have taken on the challenge of data collection and analysis. This book was written to guide researchers—inside and outside of departments—in this endeavor. It explains how social science principles can and should be applied to the analyses of vehicle stop data. It also discusses examples of the work being conducted around the country by top social scientists.

We expect that some frustration will be generated by our message that data collection cannot provide unequivocal answers to questions about the existence of racial bias by police in a jurisdiction. Despite the sincerity of most people posing the questions, answers that are definitive cannot be offered. Data analysis is not as easy as comparing stop data to jurisdiction-level census data, although police departments and concerned residents may well wish it were. We hope, however, that the frustrations that may be experienced are offset somewhat by concrete and useful advice. This book provides previously lacking specific information concerning how data can be analyzed and the results reported responsibly. We also hope frustrations are offset by the knowledge that even equivocal data can provide guidance for useful changes in a jurisdiction. A key value of these data is their potential to bring police and residents of the community together around a table to identify what might be done to make progress in the jurisdiction on the issues of racially biased policing and the perceptions of its practice.

Appendix A

USING CENSUS DATA FOR BENCHMARKING

Appendix A, written by Karen Parker of the University of Florida, has three parts: A.1, Summary Files Available from the 2000 Decennial Census; A.2, Definitions of the Geographic Units of Analysis Used by the 2000 Decennial Census and Summary Files with Data for Each Unit; and A.3, A Step-by-Step Guide to Accessing and Downloading Data from the U.S. Census Bureau Web Site.

APPENDIX A.1
SUMMARY FILES AVAILABLE FROM THE 2000 DECENNIAL CENSUS

For the four "summary files" available from the U.S. Census Bureau, we provide (1) a description by the Census Bureau of each file and (2) additional information that is relevant to the analyst conducting benchmarking with adjusted census data. Although all four files are described, note that it is Summary File 1 that will be of most value to the analyst.

Summary File 1

Summary File 1 (SF 1) contains the 100-percent data, which is the information compiled from the questions asked of all people and about every housing unit. Population items include sex, age, race, Hispanic or Latino, household relationship, and group quarters. Housing items include occupancy status, vacancy status, and tenure (owner occupied or renter occupied).

There is a total of 171 population tables (identified with a "P") and 56 housing tables (identified with an "H") shown down to the block level, and 59 population tables shown down to the census tract level (identified with a "PCT") for a total of 286 tables. There are 14 population tables and 4 housing tables shown down to the block level, and 4 population tables shown down to the census tract level that are repeated by major race and Hispanic or Latino groups.

The major race and Hispanic or Latino groups are: White alone; Black or African American alone; American Indian and Alaska Native alone; Asian alone; Native Hawaiian and Other Pacific Islander alone; Some other race alone; Two or more races; Hispanic or Latino; and White alone, not Hispanic or Latino.

SF 1 includes population and housing characteristics for the total population, population totals for an extensive list of race (American Indian and Alaska Native tribes, Asian, and Native Hawaiian and Other Pacific Islander) and Hispanic or Latino groups, and population and housing characteristics for a limited list of race and Hispanic or Latino groups. Population and housing items may be cross tabulated. Selected aggregates

> and medians also are provided. A complete listing of subjects in this file is found in the section, "Subject Locator."
>
> Source: U.S. Census Bureau, 2000 Census of Population and Housing, Summary File 1: Technical Documentation, 2001, p. 1-1. See www.census.gov.

As indicated above in the official description of SF 1 by the Census Bureau, this summary file contains basic tabulations of information collected on all people and housing units. It includes counts for many detailed race and Hispanic or Latino categories. This file will be the primary, if not sole, source of data for the analyst conducting adjusted census benchmarking.

Some (not all) of the information available in SF 1 is listed below. For a complete list of all person, household, family and housing unit characteristics, see "Data Sets" at http://factfinder.census.gov.

Person Characteristics:

- Total Population
- Urban and Rural
- Race
- Hispanic or Latino
  - By Race
- Race for the population 18 years and over
- Hispanic or Latino
  - By Race for the population 18 years and over
- Sex
  - By Age
  - By Race
  - By Hispanic or Latino
- Median Age
  - By Sex
  - By Race
  - By Hispanic or Latino
- Households
  - By Race
  - By Hispanic or Latino

Household Characteristics:
- Population in Households
  - By Race
  - By Hispanic or Latino
- Average Household Size
  - By Race
  - By Hispanic or Latino

Family Characteristics:
- Families
  - By Race
  - By Hispanic or Latino
- Population in Families
  - By Race
  - By Hispanic or Latino
- Average Family Size
  - By Race
  - By Hispanic or Latino

Housing Unit Characteristics:
- Housing Units
- Urban and Rural
- Occupancy Status
- Tenure
- Vacancy Status
- Race of Householder
  - By Hispanic or Latino
- Total Population in Occupied Housing Units
  - By Tenure
  - By Race
  - By Hispanic or Latino
- Average Household Size of Occupied Housing Units
  - By Tenure
  - By Race
  - By Hispanic or Latino

Summary File 2

Summary File 2 (SF 2) contains the 100-percent data (the information compiled from the questions asked of all people and about every housing unit). Population items include sex, age, race, Hispanic or Latino, household relationship, and group quarters. Housing items include occupancy status, vacancy status, and tenure (owner occupied or renter occupied).

SF 2 includes population characteristics, such as sex by age, average household size, household type, relationship by household type (including living alone), unmarried-partner households, nonrelatives by household type, and own children under 18 years by family type and age. The file includes housing characteristics, such as tenure, tenure by age of householder, and tenure by household size for occupied housing units. Selected aggregates and medians also are provided. . . .

These 100-percent data are presented in 36 population tables (matrices) and 11 housing tables, identified with “PCT” and “HCT,” respectively. Each table is iterated for 250 population groups: the total population, 132 race groups, 78 American Indian and Alaska Native tribe categories (reflecting 39 individual tribes), and 39 Hispanic or Latino groups. The presentation of SF 2 tables for any of the 250 population groups is subject to a population threshold of 100 or more people. That is, if there are fewer than 100 people in a specific population group in a specific geographic area, their population and housing characteristics data are not available for that geographic area in SF 2. . . .

Source: U.S. Census Bureau, 2000 Census of Population and Housing, Summary File 2: Technical Documentation, 2001, p. 1-1. See www.census.gov.

Summary File 2 will be of limited use to analysts conducting census benchmarking. It contains information on many of the same variables included in SF 1, but the presentation of race/ethnicity data in SF 1 is superior to that in SF 2.

Some (not all) of the subjects covered in SF 2 are listed below. For a complete listing, see “Data Sets” at http://factfinder.census.gov.

Person Characteristics:
- Total Population
- Urban and Rural
- Sex By Age
- Median Age By Sex
- Households

Household Characteristics:
- Population in Households
- Average Household Size

Family Characteristics:
- Families
- Population in Families
- Average Family Size

Housing Unit Characteristics:
- Housing Units
- Urban and Rural
- Occupancy Status
- Tenure
- Vacancy Status
- Total Population in Occupied Housing Units
- Average Household Size of Occupied Housing Units

Summary File 3

Summary File 3 (SF 3) contains the sample data, which is the information compiled from the questions asked of a sample of all people and housing units. Population items include basic population totals; urban and rural; households and families; marital status; grandparents as caregivers; language and ability to speak English; ancestry; place of birth, citizenship status, and year of entry; migration; place of work; journey to work (commuting); school enrollment and educational attainment; veteran status; disability; employment status; industry, occupation, and class of worker; income; and poverty status. Housing items include basic housing totals; urban and rural; number of rooms; number of bedrooms; year moved into unit; household size and occupants per room; units in structure; year structure built; heating fuel; telephone service; plumbing and kitchen facilities; vehicles available; value of home; monthly rent; and shelter costs.

In Summary File 3, population tables are identified with a "P" and housing tables are identified with an "H" prefix, followed by a sequential number. The "P" and "H" tables are shown for the block group and higher levels of geography, while the "PCT" and "HCT" tables are shown for the census tract and higher levels of geography. There are 16 "P" tables, 15 "PCT" tables, and 20 "HCT" tables that bear an alphabetic suffix on the table number, indicating that they are repeated for nine major race and Hispanic or Latino groups.

The major race and Hispanic or Latino groups are: White alone; Black or African American alone; American Indian and Alaska Native alone; Asian alone; Native Hawaiian and Other Pacific Islander alone; Some other race alone; Two or more races; Hispanic or Latino; and White alone, not Hispanic or Latino.

Summary File 3 contains a total of 813 unique tables—484 population tables and 329 housing tables. SF 3 includes population and housing characteristics for the total population and for a limited list of race and Hispanic or Latino groups. Population and housing items may be cross tabulated. Selected aggregates and medians also are provided. . . .

Source: U.S. Census Bureau, 2000 Census of Population and Housing, Summary File 3: Technical Documentation, 2002, pp. 1-1 and 1-2. See www.census.gov.

Summary File 3 contains race, ethnicity, age, and gender information for some variables that are not included in SF 1, but these variables are not needed by the researcher benchmarking stop data. SF 3 contains data on social, economic, and housing characteristics compiled from a sample of approximately 19 million housing units (about 1 in 6 households) that received the Census 2000 long-form questionnaire.

Some information is repeated for these nine (race and Hispanic or Latino) groups: White alone; Black or African

American alone; American Indian and Alaska Native alone; Asian alone; Native Hawaiian and Other Pacific Islander alone; Some other race alone; Two or more races; Hispanic or Latino; and White alone, not Hispanic or Latino. Some information is repeated by sex (male/female), age groups, and/or a combination of these characteristics (for example, by sex and age).

Some (not all) of the available information is listed below. For a complete listing, see "Data Sets" at http://factfinder.census.gov.

Social Characteristics:
- Ancestry
- Citizenship Status
- Disability Status
  - By Age
  - By Sex
- Education Attainment (persons age 25 and older)
  - By Sex
  - By Race
  - By Hispanic or Latino
- Grandparents as Caregivers
- Households and Families
  - By Age
  - By Race
  - By Hispanic or Latino
- Language and Ability to Speak English
  - By Age
- Marital Status (persons age 15 and older)
  - By Age
  - By Sex
- Migration
  - By Race
  - By Hispanic or Latino
- Nativity and Place of Birth
  - By Race
  - By Hispanic or Latino

- Region of Birth of Foreign Born
- School Enrollment
  - By Age
  - By Sex
  - By Race
  - By Hispanic or Latino
- Urban and Rural
- Veteran Status (persons age 18 and older)
  - By Age
  - By Sex
  - By Race
  - By Hispanic or Latino

Economic Characteristics:

- Class of Worker
  - By Age
  - By Sex
- Employment Status (persons age 16 and older)
  - By Sex
  - By Race
  - By Hispanic or Latino
- Commuting to Work
- Income (persons age 16 and older)
  - By Sex
  - By Race
  - By Hispanic or Latino
- Industry (persons age 16 and older)
  - By Sex
  - By Race
  - By Hispanic or Latino
- Occupation (persons age 16 and older)
  - By Sex
- Poverty Status
  - By Age
  - By Sex
  - By Race
  - By Hispanic or Latino

Housing Characteristics:

- Journey to Work (persons age 16 and older)
- Heating Fuel
- Household Size
- Occupants per Room
- Monthly Rent
- Number of Bedrooms
- Number of Rooms
- Plumbing and Kitchen Facilities
- Telephone Service
- Units in Structure
- Value of Home
- Vehicles Available (persons aged 16 plus)
- Year Householder Moved into Unit
- Year Structure Built

Summary File 4

Summary File 4 (SF 4) contains the sample data, which is the information compiled from the questions asked of a sample of all people and housing units. Population items include basic population totals; urban and rural; households and families; marital status; grandparents as caregivers; language and ability to speak English; ancestry; place of birth, citizenship status, and year of entry; migration; place of work; journey to work (commuting); school enrollment and educational attainment; veteran status; disability; employment status; industry, occupation, and class of worker; income; and poverty status. Housing items include basic housing totals; urban and rural; number of rooms; number of bedrooms; year moved into unit; household size and occupants per room; units in structure; year structure built; heating fuel; telephone service; plumbing and kitchen facilities; vehicles available; value of home; monthly rent; and shelter costs.

In Summary File 4, the sample data are presented in 213 population tables (matrices) and 110 housing tables, identified with "PCT" and "HCT," respectively. Each table is iterated for 336 population groups: the total population, 132 race groups, 78 American Indian and Alaska Native tribe categories (reflecting 39 individual

> tribes), 39 Hispanic or Latino groups, and 86 ancestry groups. The presentation of SF 4 tables for any of the 336 population groups is subject to a population threshold. That is, if there are fewer than 100 people (100 percent count) in a specific population group in a specific geographic area, and there are fewer than 50 unweighted cases, their population and housing characteristics data are not available for that geographic area in SF 4. . . .
>
> Population and housing items may be cross tabulated. Selected aggregates and medians also are provided. . . .
>
> Source: U.S. Census Bureau, 2000 Census of Population and Housing, Summary File 4: Technical Documentation, 2003, p. 1-1. See www.census.gov.

Summary File 4 is of limited value for adjusted census benchmarking. Like SF 3, it presents information on the population and housing data collected on a sample basis from the Census 2000. SF 4 is repeated or iterated for the total population and 335 additional population groups: 132 race groups, 78 American Indian and Alaska Native tribe categories, 39 Hispanic or Latino groups, and 86 ancestry groups.

Population and housing data for any of the above population groups will be shown only if there are at least 50 unweighted sample cases in a specific geographic area. This file presents data on the population and housing known as the "Sample Data" because they are obtained from questions asked of a sample (generally 1-in-6) of persons and housing units.

Some information is repeated by sex (male/female), age groups, and/or a combination of other characteristics (for example, by sex and age, by sex by age by place of birth). Some (not all) of the information in Summary File 4 is listed below. For a complete listing, see "Data Sets" at http://factfinder.census.gov.

Person Characteristics:
- Ancestry
  - By Sex
  - By Age

- Citizenship Status
    - By Sex
    - By Age
    - By Place of Birth
- Disability
    - By Sex
    - By Age
    - By Employment Status
- Educational Attainment
- Employment Status
- Grandparents as Caregivers
- Households and Families
- Income in 1999
- Industry, Occupation, and Class of Worker
    - By Sex
- Journey to Work (commuting)
- Language and Ability to Speak English
- Marital Status
- Migration
- Place of Birth
- Place of Work
- Poverty Status in 1999
    - By Sex
    - By Age
    - By Educational Attainment
    - By Public Assistance
    - By Place of Birth
- School Enrollment
    - By Sex
    - By Age
- Veteran Status
- Work Status in 1999
- Year of Entry
    - By Sex
    - By Place of Birth

Housing Characteristics:

- Bedrooms
  - By Rent
- Heating Fuel
- Kitchen Facilities
- Mortgage Status
- Plumbing Facilities
  - By Age of Householder
  - By Tenure
- Real Estate Taxes
- Rooms
- Selected Monthly Owner Costs (utilities, insurance, fuel costs)
- Telephone Services
- Units in Structure
  - By Age of Householder
  - By Household Income
- Value of Home or Monthly Rent Paid
  - By Occupied Housing Status
  - By Tenure
- Vehicles Available
  - By Tenure
- Year Moved into Structure
- Year Structure Built
  - By Tenure
  - By Age of Householder

Subjects included in SF 4 but also covered in Summary Files 1, 2, and 3 are:

- Age
- Hispanic or Latino Origin
- Household Relationship
- Race
- Sex
- Tenure
- Vacancy Status

APPENDIX A.2
DEFINITIONS OF THE GEOGRAPHIC UNITS OF ANALYSIS USED BY THE 2000 DECENNIAL CENSUS AND SUMMARY FILES WITH DATA FOR EACH UNIT

Area name	Area definition	SF 1	SF 2	SF 3	SF 4
Nation	U.S. geographical boundary.	Y	Y	Y	Y
Region	Four groupings of states (Northeast, South, Midwest, and West).	Y	Y	Y	Y
Division	A grouping of states within a geographic region. Currently the census has defined nine divisions.	Y	Y	Y	Y
State	The primary legal subdivision of the United States.	Y	Y	Y	Y
Block	A subdivision of a census tract. A block is the smallest geographic unit for which the Census Bureau provides tabular data. Many blocks correspond to individual city blocks bounded by streets. Especially in rural areas, blocks may include many square miles and may have some boundaries that are not streets. Blocks are defined uniquely within a census tract by a four-digit number.	Y			
Block Group	A subdivision of a census tract. A block group consists of all the blocks within a census tract with the same beginning number.	Y		Y	
ZIP Codes	A ZIP Code tabulation area is a statistical geographic entity that approximates the delivery area for a U.S. Postal Service five-digit or three-digit ZIP Code.	Y		Y	
Census Tract	A small, relatively permanent statistical subdivision of a county. Census tract boundaries are always nested within counties and designed to be relatively homogeneous units with respect to characteristics; census tracts average about 4,000 inhabitants.	Y	Y	Y	Y

Central City	The largest city and, in some cases, one or more additional cities in a metropolitan area (MA). In a number of instances, only part of a city qualifies as central, because another part of the city extends beyond the MA boundary.	Y	Y	Y	Y
City/ Consolidated	A type of incorporated place in 49 states and the District of Columbia in which the functions of the place and its county or minor civil division have merged.	Y	Y	Y	Y
County	The primary legal subdivision in most states. In Louisiana, these subdivisions are known as parishes. In Alaska, which has no counties, the county equivalents are boroughs. In four states (Maryland, Missouri, Nevada and Virginia), there are one or more cities that are independent of any county and thus constitute primary subdivisions of the state.	Y	Y	Y	Y
Place	A concentration of population either legally defined as an incorporated place or defined for statistical purposes as a census designated place. Typically used by most researchers to identify "city" boundaries.	Y	Y	Y	Y

Source: U.S. Census Bureau, 2000 Census of Population and Housing, Geographic Terms and Concepts. For additional information, see http://www.census.gov/geo/www/reference.html.

APPENDIX A.3
STEP-BY-STEP GUIDE TO ACCESSING AND DOWNLOADING DATA FROM THE U.S. CENSUS WEB SITE

We explain step-by-step how researchers can accomplish four tasks: locating geographic areas or subareas of a jurisdiction, obtaining race and ethnicity information for various ages, obtaining information on vehicle-less households by race/ethnicity, and obtaining average household size by race/ethnicity.

Locating Geographic Areas or Subareas of a Jurisdiction

To help researchers learn how to access the census data they need, we present here the steps to follow for one particular example: obtaining vacancy status information for housing units in census tract 2.03 at the block group level in Miami-Dade County, Florida.

Step 1: Go to http://factfinder.census.gov

Step 2: Under "Data Sets," select the summary file (for instance, SF 1) that contains information you wish to obtain (see Appendix A.1 and Appendix A.2).

Because the vacancy status is available in Summary File 1 at the block group level, select "2000 Summary File 1" of the decennial census. From the options at the right side of the page, choose the "about this data set" option.

Step 3: To make a data request, click on "detailed tables." This option will allow you to specify the parameters and geographic areas of interest for the information you need.

At this point, factfinder provides a series of drop-down options.

Step 4: Leave "Choose a Selection Method" in the default option, which is "list." Specify the "geographic type" from the available drop-down list that includes “state,” “census tract,” “block group,” and “block,” among others. Choose “block group.”

Factfinder will prompt the researcher to select the “state” of interest. Choose “Florida.”

Factfinder will now prompt the researcher to enter in the “county” where the block groups of interest are located. In the drop-down window, choose “Miami-Dade.”

Then specify census tract 2.03.

Step 5: Choose “all block groups” from the next drop-down window and then click “add.”

Step 6: Once all the block groups appear in the bottom window, click “next,” which will allow the researcher to specify the data tables of interest.

At the top of the new page, leave the “search” option in the default position, which is “show all tables.” All the data tables available in Summary File 1 for the geographic type specified in earlier steps will appear in the next window. Scroll through the table options to find the table of interest (vacancy status).

Highlight “H5: Vacancy Status (Vacant Housing units)” and then “add” this highlighted table to the next window.

Step 7: Once all tables of interest are highlighted and listed in the bottom window, select “show table” by single clicking on the prompt.

A data table will appear that looks like Table A.3.1.

Step 8: Note that in the upper righthand corner of the screen, Factfinder gives the researcher the options to print or download the data table. If the researcher chooses to print, the information displayed on the screen will print as shown. If the researcher chooses to download, two options are available: download the tables in presentation ready format or save the tables into a data base**.** **Presentation ready format** preserves the

Table A.3.1. Vacancy Status by Housing Type for All Block Groups in Census Tract 2.03 in Miami-Dade County, Florida

	Block Group 1, Census Tract 2.03, Miami-Dade County, Florida	Block Group 2, Census Tract 2.03, Miami-Dade County, Florida	Block Group 3, Census Tract 2.03, Miami-Dade County, Florida	Block Group 4, Census Tract 2.03, Miami-Dade County, Florida	Block Group 4, Census Tract 2.03, Miami-Dade County, Florida	Block Group 6, Census Tract 2.03, Miami-Dade County, Florida
Total:	8	446	13	100	16	10
For rent	0	79	1	16	1	0
For sale only	3	27	4	9	4	4
Rented or sold, not occupied	3	27	3	9	3	1
For seasonal, recreational, or occasional use	2	278	3	59	1	4
For migrant workers	0	0	0	0	0	0
Other vacant	0	35	2	7	7	1

Source: U.S. Census Bureau, 2000 Census.

table format, title, head note, and footnote(s) exactly as shown on the screen. Use these file formats if you need to insert tables directly into other documents. In the **data base ready format,** *only* data rows are downloaded. The table format, title, head note or footnote(s) are *excluded* from the download. Use the data base file if you intend to manipulate the data.

Presentation ready format options include:

1. Comma delimited
2. Tab delimited
3. Rich text format

Comma delimited and Tab delimited file formats are .txt files. The **rich text file format** is a word processor ready format (.rtf). That is, this file type allows you to open the table in any word processor (Word, WP, etc.). You also have the option to transpose rows and columns using any of the three options.

NOTE: The census bureau recommends the rich text file format over the other options when downloading the tables. If the researcher's browser recognizes the .rtf format, it will open the file automatically in the same window.

Data base ready formats include:

1. Microsoft Excel – This is the spreadsheet ready file format (.xls file).
2. Comma delimited database **–** This file format is for downloading the data records in order to load them into database software for data manipulation (.txt file).

All data base ready download files are compressed into one file named **output.zip**. This compressed file contains:

- One or more **data file(s)** – The naming format and number of these files will vary by data set and/or by the number of tables you have selected.
- One **geographic identifier file** – This file allows you to link multiple data files.

- One **readme.txt file –** This file explains the naming and content of the downloaded files and how to link them together.
- One **data set specific readme.txt file** – This file explains any data set anomalies and the location of the specific technical documentation for the data set

After you make your selections, click "OK." Now save your file in the location of your choice. That is, you will have the option to save to diskette, save to your computer hard drive, or open the table in the format you requested (for example, Excel, Word, etc.).

If you want to produce additional tables, change to a different summary file (SF 1-4) or choose additional geographical areas, select "change selections" at the top of the screen.

Obtaining Race and Ethnicity Information for Various Ages

We present here the steps to follow for one particular example: obtaining the number of Black, White, and Hispanic residents between 15 and 24 years of age for all census tracts in Carroll County, Indiana. As explained in Chapter 5, researchers will need to determine if the proportion of residents between the legal driving age and 24 within each racial/ethnic group are equivalent. They will also need to determine the race/ethnic breakdown of the residential population for residents over the driving age to produce their census benchmark.

Steps 1–3: Follow the steps explained above. Factfinder will then prompt the researcher to select the geographical areas of interest through a series of drop-down options.

Steps 4–6: Specify the "geographic type" as "census tract," "select a state" as "Indiana" from the drop-down list, then "select a county" as "Carroll County." The next drop-down box (that is, "select one or more geographical areas and then click

'ADD'") will allow the researcher to add "all census tracts" from the list because all tracts located in Carroll County are of interest. Once all the census tracts are listed in the bottom box, click "next."

At the top of the new page, leave the "search" option in the default position, which is "show all tables." All the data tables available in summary file (SF) 1 for the geographic type specified will appear in the next window. Scroll through the table options to find the table of interest (age information for race and ethnic groups).

After scrolling through the tables, notice that the census provides detailed information on residents' age by sex in table "P12: Sex by Age (total population)." Furthermore, the Census Bureau provides a breakdown of sex by age information for each race and ethnic group in tables P12A-P12I.

Highlight "P12A: Sex by Age (White alone)," "P12B: Sex and Age (Black Alone)" and "P12H: Sex by Age (Hispanic or Latino)." Add each table to the next window.

Step 7: Once all tables of interested are highlighted and listed in the bottom window, select "show table" by single clicking on the prompt.

The three tables—P12A, P12B, and P12H—will appear. Only P12A is reproduced below. It shows population counts of "Whites" within the seven census tracts of Carroll County. P12B and P12H (not shown) convey the same information for people who identify as "Black or African American" and as "Hispanic or Latino," respectively.

Step 8: Print the tables as they appear on the screen or download the data for statistical manipulation or computational use. (See detailed instructions in Step 8 of the first example.)

Using this information provided in the tables, determine the number of White, Black, and Hispanic residents between the ages of 15 and 24 for each census tract located in Carroll County, Indiana. To obtain this information, add the different

Table P12A. Sex by Age (White Males) continued from previous page

	Census Tract 9593, Carroll County, IN	Census Tract 9594, Carroll County, IN	Census Tract 9595, Carroll County, IN	Census Tract 9596, Carroll County, IN	Census Tract 9597, Carroll County, IN	Census Tract 9598, Carroll County, IN	Census Tract 9599, Carroll County, IN
Total:	2,958	2,453	2,468	3,177	3,274	2,923	2,438
Male:	1,464	1,253	1,268	1,533	1,662	1,371	1,244
Under 5 years	95	69	54	110	141	90	90
5 to 9 years	122	118	79	118	133	106	90
10 to 14 years	119	105	85	134	140	119	100
15 to 17 years	67	56	38	68	80	62	62
18 and 19 years	24	31	32	40	50	42	36
20 years	4	15	13	15	18	12	11
21 years	12	10	16	15	17	9	6
22 to 24 years	44	40	28	49	55	48	32
25 to 29 years	90	79	72	102	117	90	61
30 to 34 years	90	82	74	130	107	84	83
35 to 39 years	140	100	101	107	121	107	96
40 to 44 years	136	118	100	122	138	115	107
45 to 49 years	102	120	115	114	137	85	94
50 to 54 years	93	91	101	76	121	79	92
55 to 59 years	76	75	97	79	78	76	75
60 and 61 years	28	15	34	22	19	22	29
62 to 64 years	46	24	49	40	36	24	36
65 and 66 years	11	17	28	19	15	16	24
67 to 69 years	30	16	43	27	27	22	30
70 to 74 years	46	36	47	52	46	53	37
75 to 79 years	39	25	37	46	30	54	21
80 to 84 years	31	10	15	28	25	33	20
85 years and over	19	1	10	20	11	23	12

Source: U.S. Census Bureau, 2000 Decennial Census. Data set: Census 2000 Summary File 1.

Table P12A. Sex by Age (White Females) continued on next page

	Census Tract 9593, Carroll County, IN	Census Tract 9594, Carroll County, IN	Census Tract 9595, Carroll County, IN	Census Tract 9596, Carroll County, IN	Census Tract 9597, Carroll County, IN	Census Tract 9598, Carroll County, IN	Census Tract 9599, Carroll County, IN
Total:	2,958	2,453	2,468	3,177	3,274	2,923	2,438
Female:	1,494	1,200	1,200	1,644	1,612	1,552	1,194
Under 5 years	111	70	72	105	131	97	61
5 to 9 years	100	77	69	108	128	106	82
10 to 14 years	115	107	63	107	140	101	90
15 to 17 years	67	57	44	77	84	57	56
18 and 19 years	30	36	22	38	52	31	24
20 years	13	8	12	14	19	20	15
21 years	8	4	15	9	13	18	8
22 to 24 years	43	35	42	53	53	47	32
25 to 29 years	106	71	59	110	94	85	57
30 to 34 years	87	81	73	108	107	94	80
35 to 39 years	127	97	85	121	131	94	98
40 to 44 years	112	117	106	109	129	111	106
45 to 49 years	92	108	89	95	122	113	84
50 to 54 years	99	80	107	80	98	65	88
55 to 59 years	78	62	85	75	72	78	79
60 and 61 years	24	19	40	20	24	26	35
62 to 64 years	35	30	51	43	39	36	40
65 and 66 years	18	16	29	30	18	35	18
67 to 69 years	25	26	30	36	46	33	26
70 to 74 years	75	49	45	84	41	83	40
75 to 79 years	57	25	29	87	32	87	34
80 to 84 years	45	11	25	79	22	66	25
85 years and over	27	14	8	56	17	69	16

Source: U.S. Census Bureau, 2000 Decennial Census. Data set: Census 2000 Summary File 1.

age categories together (15 to 17, 18 to 19, 20, 21, 22 to 24) for males and then for females. Combine the age information for males and females to get the total number of residents between the ages of 15 and 24 per census tract location. This step will be repeated for each race and ethnic group.

Obtaining Information on Vehicle-less Households by Race/Ethnicity

Chapter 5 describes how to adjust census data to account for households that have no vehicles. Here we describe where to locate relevant information to implement this adjustment.

Summary File 3, tables HCT33 A-I, shows the number of occupied housing units with "no vehicle available" and "1 or more vehicles available" for a selected geographic area (census tract) by race and ethnicity.

For purposes of illustration, we display this table for one racial group (White residents) at the census tract level (within Miami-Dade County). Here's the tabular information using the rich text file format option described above.

Table HCT33A. Vehicles Available (White Alone Householder)

	Census Tract 1.06, Miami-Dade County, Florida	Census Tract 1.08, Miami-Dade County, Florida	Census Tract 1.09, Miami-Dade County, Florida	Census Tract 1.10, Miami-Dade County, Florida
Total:	2,491	1,780	511	2,688
No vehicle available	203	357	74	213
1 or more vehicles available	2,288	1,423	437	2,475

Source: U.S. Census Bureau, 2000 Decennial Census. Data set: Census 2000 Summary File 3.

Average Household Size by Race/Ethnicity

As indicated in Chapter 5, a researcher who is adjusting census data for household vehicle ownership will need to transform the household-level information into individual-level information using information regarding the average number of individuals per household for each racial/ethnic group. This information is available in Summary File 1 in tables for "Average Household Size" that can be produced for each racial/ethnic group. For instance, see P17B "Average Household Size (Black or African American Alone Households)."

Appendix B

APPENDIX B.
TRANSFORMING TWO-VARIABLE CENSUS DATA INTO A SINGLE RACE/ETHNICITY VARIABLE

If a law enforcement agency's form for recording police-citizen contacts includes Hispanic within a single race/ethnicity category, its researchers will need to transform the census data on jurisdiction residents from a two-variable structure into a single-variable structure. The hypothetical data presented in Appendix Table B will help explain how.

As noted in Chapter 5, the U.S. Census Bureau treats race and Hispanic Origin (referred to here as "ethnicity") separately. Appendix Table B, Panel 1, presents the number of jurisdiction residents by the separate variables of race and ethnicity. The census also provides information on the combined race/ethnicity of jurisdiction residents (Panel 2). In this panel the different races in the jurisdiction are presented by Hispanic origin or Non-Hispanic origin.

Adjusted census benchmarking requires that the census data and law enforcement agency data be comparable in structure. The result of the transformation to accomplish this is shown in Panel 3. First, jurisdiction residents that self-identify as being of Hispanic origin in Panel 2 would be subtracted from their respective race categories in Panel 1 to produce a new, lower tally for each race. Second, the 25,000 Caucasian-Hispanics, 3,000 African American-Hispanics, 600 Asian-Hispanics, and 50 Other Race-Hispanics would be added to produce a new "Hispanics" category shown in Panel 3.

Table B: Using Census Information Shown in Panels 1 and 2 to Produce Transformed Data in Panel 3, Hypothetical Data

PANEL 1	
Race	**Number**
Caucasian	75,000
African American	13,000
Asian	3,600
Other	300
TOTAL	91,900
Ethnicity	**Number**
Hispanic Origin	28,650
Non-Hispanic Origin	63,250
TOTAL	91,900
PANEL 2	
Race by Ethnicity	**Number**
Caucasian, Non-Hispanic	50,000
Caucasian, Hispanic	25,000
African American, Non-Hispanic	10,000
African American, Hispanic	3,000
Asian, Non-Hispanic	3,000
Asian, Hispanic	600
Other, Non-Hispanic	250
Other, Hispanic	50
TOTAL	91,900
PANEL 3	
Race/Ethnicity	**Number**
Caucasian	50,000
African American	10,000
Asian	3,000
Other	250
Hispanic	28,650
TOTAL	91,900

Appendix C

APPENDIX C
TRANSFORMING AGENCY DATA OR DMV DATA TO PRODUCE COMPARABLE MEASURES OF RACE AND ETHNICITY

Benchmarking to assess racially biased policing in a jurisdiction requires comparable stop data and benchmark data. In several benchmarking methods (for example, benchmarking with DMV data, described in Chapter 6, and benchmarking with data from "blind" enforcement mechanisms, described in Chapter 7), law enforcement agencies compare stop data and data from a state's Department of Motor Vehicles. Stop data and DMV data can vary in six ways. In four of them, a transformation is possible to make the data comparable:

1. If the state's DMV provides information on race alone and not on ethnicity, and if race and ethnicity are treated separately on the law enforcement agency's police-citizen contact data form, then analyses can be conducted to assess potential bias based on race only. The law enforcement agency would compare the racial profile of drivers stopped by police to the racial profile of people with a driver's license. The separate ethnicity variable must be ignored because no benchmark for it exists.

2. Conversely, if the DMV in the state has separate race and ethnicity variables but the agency form requests information on race only, the analyst would have to ignore the DMV's ethnicity information because she or he would not have the corresponding information in the stop data.
3. If the DMV has race/ethnicity in one variable, and if the agency's stop forms have race and ethnicity as separate variables, the analyst would transform the stop data to match the DMV data.[1] Appendix B explains how this can be accomplished. Note that in the transformation explained in Appendix B, the race by ethnicity information was available in the census data. In benchmarking that relies on DMV data with a single race/ethnicity variable, the race by ethnicity information is available in the stop data.
4. If the DMV has race/ethnicity combined into one variable, and if the agency's stop forms request information on race only, the analyst can use U.S. Census information to estimate the race data for the Hispanics in the data set. That is, the analyst can determine for the jurisdiction population of driving age the races of the Hispanic population based on the census data and then use that information to estimate the races of the Hispanics with a driver's license. For instance, if the census data indicate that 20 percent of the Hispanics of driving age are Caucasian, the analyst can reasonably estimate that 20 percent of the Hispanics with a driver's license are Caucasian, and so forth for each race category.

The appendix table below describes the four combinations of race and ethnicity measurements we have presented thus far and the transformations required. It also describes other possi-

---

[1] With census benchmarking, the opposite occurs: the census data are transformed to match the stop data.

ble combinations. In two cases, as the table shows, the incompatibility of the DMV data and the agency data cannot be overcome unless Hispanics comprise a very small percentage (for example, less than 5 percent) of the people stopped by police and of the people in the jurisdiction with a driver's license.

Appendix C Table
Measures of Race and/or Ethnicity, by DMV Data and Stop Data

STOP DATA	DMV Data on Registered Vehicle Owner		
	VARIABLE FOR RACE ONLY	SEPARATE VARIABLES FOR RACE AND ETHNICITY	RACE/ETHNICITY COMBINED IN ONE VARIABLE
VARIABLE FOR RACE ONLY	Measurements are matched; can proceed.	Can analyze data using race information only.	Can produce single race variable, using census data, by estimating the race of Hispanics who have a driver's license.
SEPARATE VARIABLES FOR RACE AND ETHNICITY	Can analyze data using race information only.	Measurements are matched; can proceed.	Transform separate variables in the stop data into single combined variable.
RACE/ETHNICITY COMBINED IN ONE VARIABLE	Cannot proceed with this method unless Hispanic groups in both the stop data and DMV data are small.	Cannot proceed with this method unless Hispanic groups in both the stop data and DMV data are small.	Measurements are matched; can proceed.

5. If the DMV has information only on race, and if the agency includes ethnicity in a single race/ethnicity variable on its form, the agency cannot use this benchmarking method. The only exception is if the Hispanic population is very small in both data sets. Then the analyst would exclude Hispanics from the numerator (stop data) and calculate the racial profile using only the remaining stops.
6. If the DMV has race and ethnicity separated into two variables, and if the agency has race and ethnicity combined into one variable, the agency also cannot use this benchmarking method. Again, the only exception is if the Hispanic population is very small in both data sets. Then the analyst would exclude Hispanics from the numerator (stop data) and benchmark the race of people stopped against the race of people with a driver's license; the analyst would ignore the DMV information on ethnicity.

In situations 5 and 6, the jurisdiction is, quite unfortunately, reducing the scope of its assessment of racially biased policing; the analyst is able to test only for racial bias and not for bias based on ethnicity.

Appendix D

APPENDIX D
MAKING THE CASE FOR MEASURING "WHO IS DRIVING" INSTEAD OF "WHO IS VIOLATING"

by John Lamberth, David Harris, Jack McDevitt, and Deborah Ramirez

One question facing those attempting to analyze traffic stop data involves the selection of the most appropriate benchmark to use for comparison. A number of measures have been used in the research to date and an open question remains as to whether using estimates of the population violating traffic laws (hereafter referred to as "violators") is an improvement over estimates of drivers operating on a community's roadways (hereafter referred to as "traffic"). Some early court decisions (including the *Soto* and *Wilkins* decisions)[1] originally held that the appropriate benchmark was a profile of violators, but then quickly changed their focus when it became obvious that the two groups–violators and traffic–were virtually synonymous populations. That is, these two courts and others have held that

---

1 *State v. Pedro Soto*, 734 A.2d 350 (N.J. Super. Ct. Law Div. 1996); *Wilkins v. Maryland State Police*, Settlement Agreement, Civil No. MJG-93-468 (D. Md. 1995).

an appropriate benchmark would characterize the traffic on the relevant roadways.

Court decisions uniformly support the notion that any motorist violating a traffic law is subject to being stopped by police and thus motorists are the appropriate group to use in formulating a benchmark. Empirical evidence strongly supports the contention that traffic and violators are synonymous, and in *Soto* the court essentially considered them equivalent.

In the earliest scientific attempts to develop benchmarks for police stops, the research team (headed by the first author of this piece) determined both the proportion of Black motorists in the traffic stream and among those violating at least one traffic law (*New Jersey v. Soto*, et al.).[2] That is, the team developed both profiles. The results of that analysis, and subsequent analyses, determined that the two populations are virtually synonymous. First, in the research conducted for the cases of *Soto* and in *Wilkins v. Maryland State Police* (MSP) virtually every motorist was speeding (98.3 percent in *Soto* and 93.3 percent in *Wilkins*). More recently, Lamberth[3] reported a study in which police officers were given five minutes to determine whether randomly selected cars were violating some traffic law. The study concluded that fully 94 percent of the drivers were violating some law, and it took a mean of 28 seconds for the officers to spot the violation.

The empirical results presented above strongly support the contention that traffic and violators are essentially synonymous. Having made the case that everyone can be legally stopped, the important issue becomes which motorists are stopped and how their racial/ethnic makeup compares to those motorists driving.

---

2 Lamberth, J. (1994) Revised statistical analysis of the incidence of police stops and arrests of black drivers/travelers on the New Jersey Turnpike between interchanges 1 and 3 from the years 1988 through 1991. Report submitted in *State v. Pedro Soto*, 734A. 2d 350 (N.J. Super. Ct. Law Div. 1996).

3 Lamberth, John, "Measuring the racial/ethnic make up of traffic: The how, what and why." Paper presented at *Confronting Racial Profiling in the 21st Century: Implications for Racial Justice*. Boston, March, 2003.

We turn next to the issue of whether all violations are equally subject to enforcement by police. We consider two interrelated suggestions: (1) the police primarily stop egregious traffic violators, and (2) minority drivers are stopped more often than non-minorities because they fall in this egregious violator category.

While it is probably true that drivers engaged in the most egregious traffic violations are most likely to be stopped by the police, it remains an open question if those egregious violators vary by race of the driver. Police make the decision on which vehicles to stop for a wide variety of reasons including the impact of the stop on the other traffic proceeding along the roadway. Supporting the argument that police do not stop only egregious traffic violators is information pertaining to the proportion of stops by law enforcement that do not result in citations. Data on vehicle stops from a number of departments (for instance, Arizona Department of Public Safety, New Jersey State Police, Maryland State Police, Washtenaw County Sheriff's Department) indicate that approximately one-third to two-thirds or more stops do not result in citations. Furthermore, even though speeding is often the most cited infraction, substantially less than half of drivers who are speeding are going sufficiently over the speed limit to qualify as an egregious violator. For example, in a recent study of nine departments in Nevada,[4] which accounted for 400,000 stops, only about 35 percent of the stops recorded were considered egregious. This means that approximately two-thirds of the stops were made of drivers who were not egregiously violating traffic laws.

Secondly, those who argue that minorities are more often in the group of egregious violators have scant empirical evidence to support their position. Some researchers have suggested that Blacks violate at least some traffic laws more egregiously than non-Blacks and therefore are more likely to be included in those motorists most likely to be stopped by police. This claim, made by the state's expert in *Soto*, was soundly rejected by the Court

---

[4] McCorkle, R.C., *A.B. 500 Traffic Stop Data Collection Study*. Report submitted to Attorney General of Nevada, Jan., 2003.

because the expert could not provide empirical support for his contention. Countering this expert's argument were five troopers and a police expert who testified that Blacks and non-Blacks could not be distinguished on the basis of their driving behavior.

One study (Lange, Blackman and Johnson, 2001) suggests that there are more Black than White egregious speeders (speeding more than 15 miles over the limit) on the New Jersey Turnpike when the speed limit is 65 miles per hour, but not when the speed limit was 55 miles per hour. This study, which has been soundly criticized, does not engender great confidence because of the difficulty in obtaining data from the study to perform independent analysis to confirm (or disconfirm) the original results. One methodological limitation of the study was that the race of about a third of the motorists could not be agreed upon by two of three coders. No information has yet been obtained on what it would be for a more scientifically defensible three of three coders, but it is safe to assume that there would be more motorists whose race/ethnicity could not be determined. Furthermore, the ambiguous finding concerning the 65- versus 55-miles per hour speed limits is unexplained.

Other evidence on speeding contradicts the assertion that minorities are more likely to be egregious speeding violators. The Nevada study cited above indicates that Blacks (32 percent) and Hispanics (30 percent) are less likely to be in the egregious violator category than are other race/ethnicity groups. Over all racial/ethnic groups, 35 percent were in the egregious violator category.

Finally, if minorities were those who egregiously violated traffic laws, they would be more frequently cited than nonminority drivers. In four jurisdictions where we have the data detailed enough to make comparisons, Blacks in all four jurisdictions and Hispanics in three are cited less frequently than are nonminorities.

An important argument against trying to measure "who is violating" is the fact that there are literally hundreds of traffic violations for which a motorist can be legally stopped. These range from the serious violations (excessive speeding, running a red light, dangerously weaving in and out of traffic, etc.) to a large number that reflect relatively minor equipment violations. And, as we have seen, non-egregious violations are the ones that generate a majority of stops.

Also related to the difficulty of measuring violations is the fact that, except for speeding, most violations are subjective, either in their definition or their enforcement. As one example, consider the violation of following too closely. There are at least two different methods for making this determination. The first is measuring the distance between a motorist and the vehicle ahead of that motorist by timing it. If the following vehicle passes a stationary point in less than 1.5 to 2 seconds after the leading car passes it, some officers call it following too closely. While we do know of some officers who carry stopwatches to make this call, the violation is more often determined on a less accurate basis of counting or other estimation. The second method is to estimate the number of car lengths between the leading and following car, with the assumption being that there should be one car length for each 10 miles of speed. Therefore at 50 miles per hour, there should be five car lengths. Only officers can tell us what they actually do to "measure" following too closely.

Finally, in the realm of measurement challenges, detecting the vast majority of traffic violations for which a motorist can be stopped from a stationary point or even from a moving vehicle is either not possible or prohibitively time consuming. That is, while many of the violations are always present (equipment violations), they may not be obvious until the vehicle is observed from several angles. Stationary observations do not allow the necessary views, and moving observations can take several minutes per vehicle to see the vehicle from all angles. Furthermore, categorizing hundreds of possible violations is an insurmountable task. Most importantly, the crucial information needed is not what traffic laws motorists are violating, but which violations officers are noting and using as a basis for stopping them. To know that motorists are violating several traffic laws is unimportant for our purposes; rather, we need to know to which violations officers attend.

Officers take a large number of factors into consideration when deciding to make a traffic stop. One important consideration is the severity of the violation but officers also legitimately take into consideration such things as the traffic flow at the time, any potential dangers to the traffic by making the stop, the priorities of the agency, the officers' attitudes about traffic enforcement in general and specific violations in particular, the time of day, and weather conditions. All of these factors and others influence the decision of an individual officer to stop a particular vehicle. No road survey or other social science measurement technique can adequately model this decision-making process.

Data from police stops (i.e., the actual people that officers do stop as opposed to those they can theoretically stop) is the more appropriate data source from which to determine who does get stopped and for which violations they are stopped. The motorists that officers actually stop is a more reliable measure of officer behavior than theoretically determining the violations for which officers could stop motorists.

For all of these reasons, we argue that the appropriate data to use in determining the race/ethnicity violation matrix are the stop data from police departments in comparison to the refined estimates (benchmarks) of the driving population. From these data we know what violations draw the attention of police, and we know the violations for which drivers are stopped. And as we have noted, the vast majority of drivers are subject to being stopped due to one violation or another. It is true that traffic stop analyses should account for variations in more egregious driving behavior (e.g., speeding more than 15 miles over the posted limit) and that separate analyses for these different levels of violation should be conducted. We believe that it is not necessary or possible to develop a benchmark that adequately measures the factors that influence a police officer to stop a particular vehicle.

Appendix E

APPENDIX E
A SUMMARY OF ARGUMENTS PERTAINING TO WHETHER ANALYSTS SHOULD MEASURE "WHO IS DRIVING" OR "WHO IS VIOLATING" FOR THE BENCHMARK POPULATION

Question	Measure “Who is Driving” (Arguments from Appendix D)	Measure “Who is Violating” (Arguments from this report)
What is the appropriate benchmark?	The appropriate benchmark is the driving population. All drivers are subject to being stopped by police.	The appropriate benchmark is the violating population. Conceptually, the benchmark is the population of drivers at risk of being stopped by police, assuming no bias. Some drivers are at greater risk than others of being stopped by police for legitimate reasons. The frequency and seriousness of their violations increase their risk of being stopped. We instruct police to make stopping decisions based on driver-violating behavior; as such, we must include this factor in our studies of their decisions.
Does the population of drivers differ from the population of violators in a meaningful way?	No. The population of drivers is the same as the population of violators because most—if not all-drivers violate traffic laws to some extent. That is, all drivers are legally subject to stops by police.	Yes. While it is, indeed, likely that all drivers engage in violations, the drivers who engage in serious violations or violate most frequently are likely at greater risk of being stopped by police. This fact makes driving quality relevant in any analysis of officers’ decisions to stop drivers.

Question	Measure "Who is Driving" (Arguments from Appendix D)	Measure "Who is Violating" (Arguments from this report)
Are the egregious violators the ones stopped by police? Are the egregious violators more at risk of being stopped by police?	No. Officers do not "primarily stop egregious violators." But, yes, it may be that "drivers engaged in the most egregious violations are (the drivers) most likely to be stopped by police."	Agree with Appendix D on both points: answers are "no" and "yes," respectively. Police do not stop only egregious violators, but it is likely that the most egregious violators are most likely to be stopped by police. It is precisely because the most egregious violators are most likely to be stopped by police, that driving quality must be a factor encompassed in the research design.
What have the courts said about the benchmarks?	Some lower court decisions have held that "who is driving" is a legitimate benchmark because the "who is driving" population is essentially the same as the "who is violating" population. Again, this is because most drivers violate traffic laws to some extent.	The few decisions by lower courts have not said it was wrong to measure "who is violating," only that it is not necessary to do so. These decisions are not the last, definitive word. Other courts with other expert witnesses might conclude differently.
Should social scientists adopt court standards?	Yes. "In heavily contested litigation, there is actually more criticism (and scrutiny) of (research) results than (occurs in the context of) scientific review" (Lamberth, 2003).	No. The social science standards found acceptable by the courts are sometimes deemed unacceptable by social scientists.

Question	Measure "Who is Driving" (Arguments from Appendix D)	Measure "Who is Violating" (Arguments from this report)
Is there evidence that racial minorities violate traffic laws at a greater rate than do nonminorities?	No. There is no clear evidence that racial minorities violate traffic laws at a greater rate than nonminorities.	No. There is no clear evidence that racial minorities violate traffic laws at a greater rate than nonminorities. As Lamberth et al. indicated in Appendix D, research has not shown definitively that one group violates more frequently or more egregiously than another. Neither has research shown definitively that there are no differences between groups. It is because researchers can't rule out the possibility of differences in driving quality across racial/ethnic groups, that researchers must consider driving quality in their analysis. It may be that nonminorities violate more frequently or more egregiously than minorities. If this is true and our analyses don't encompass this possibility by considering driving quality, we will not be able to conduct a viable assessment of the existence of racially biased policing.

Question	Measure "Who is Driving" (Arguments from Appendix D)	Measure "Who is Violating" (Arguments from this report)
Which population—drivers or violators—can be measured most reliably?	The population of drivers can be measured more reliably than the population of violators. Measuring the population of drivers who are violating the many and varied traffic laws is virtually impossible. There are hundreds of violation types and most measures (except for measures of speeding) would be subjective.	Agreed: the population of all drivers can be measured more reliably than the population of all violators. Lamberth et al. are correct that researchers cannot reliably produce a profile of the drivers who violate any of the many traffic laws. To develop a benchmark based on "who is violating" requires the researcher to select specific types of violations (such as speeding) that are most amenable to measurement. The assessment that is conducted with this benchmark is arguably more valid than one that does not consider the impact of driving quality on police decisions. However, the scope of the assessment is limited—providing only a "spot check" of racially biased policing. Further, it is possible that the "spot check" does not encompass the officers, areas, or violations that manifest racially biased policing. The "costs" of this reduced scope are offset by the benefits of a more valid analysis of the factors that affect officers' stopping decisions.

References

Alpert, Geoffrey P. 2002. Personal email communication, October.

———. 2003a. Personal verbal communication, February.

———. 2003b. Personal email communication, August.

Alpert, Geoffrey P., et al. 2003. *Toward a Better Benchmark: Assessing the Utility of Not-at-Fault Traffic Crash Data in Racial Profiling Research*. Paper presented at the conference "Confronting Racial Profiling in the Twenty-first Century: Implications for Racial Justice," sponsored by the Northeastern University Institute on Race and Justice, Boston, MA, March 8-9.

Alpert Group. 2003. *Miami-Dade Racial Profiling Study*. Draft of the methods section of the report to be submitted to the Miami-Dade County (FL) Police Department.

Alschuler, Albert W. 2002. Racial Profiling and the Constitution. *University of Chicago Legal Forum*, 163.

Ayres, Ian. 2001. *Pervasive Prejudice? Unconventional Evidence of Race and Gender Discrimination*. Chicago: University of Chicago Press.

———. 2002. Outcome Tests of Racial Disparities in Police Practices. *Justice Research and Policy*, 4: 131-142.

Banks, R. Richard. 2003. Beyond Racial Profiling: Race, Policing, and the Drug War. *Stanford Law Review*, 56: 571-603.

——. n.d. *Beyond Rights, Irrationality, and Racial Profiling*. Available at http://www.law.berkeley.edu/institutes/csls/bankspaper.pdf.

Becker, Gary S. 1993. *The Economics of Discrimination*. Chicago: University of Chicago Press.

Bickman, Leonard, Debra J. Rog, and Terry E. Hedrick. 1998. Applied Research Design: A Practical Approach. In Leonard Bickman and Debra J. Rog (Eds.), *Handbook of Applied Social Research Methods* (pp. 5-37). Thousand Oaks, CA: Sage Publications.

Black, Donald J. 1980. *The Manners and Customs of the Police*. San Diego: Academic Press.

Boniface, David. 2000. *Report on the Street Surveys Carried out for the Ethnic Monitoring Committee of Watford and Three Rivers Division of Hertfordshire Constabulary*. Unpublished manuscript, Statistics Group, University of Hertfordshire.

Borooah, Vani K. 2001. Racial Bias in Police Stops and Searches: An Economic Analysis. *European Journal of Political Economy*, 17: 17-37.

——. 2002. Economic Analysis of Police Stops and Searches: A Reply. *European Journal of Political Economy*, 18: 607-608.

Braver, E. R. 2003. Race, Hispanic Origin, and Socioeconomic Status in Relation to Motor Vehicle Occupant Death Rates and Risk Factors among Adults. *Accident Analysis and Prevention*, 35(3): 295-309.

Calnon, Jennifer M. and Robin Shepard Engel. 2002. *Further Exploration of Base Rate Methodologies for Police Traffic Stops*. Paper presented at the annual meeting of the American Society of Criminology, Chicago, IL, November 15.

Campbell DeLong Resources, Inc. 1999. *Public Perceptions of Stop Decisions by Oregon Police Officers*. PowerPoint presentation prepared for the Oregon Criminal Justice Commission.

Canter, Phil. 2003. Personal phone communication with Phil Canter of the Analysis Unit of the Baltimore County Police Department, April 3.

Chakravarty, Shanti P. 2002. Economic Analysis of Police Stops and Searches: A Critique. *European Journal of Political Economy*, 18: 597-605.

Clayton, Jerry. 2004. *Enhancing Law Enforcement and Community Trust*. PowerPoint presentation at the conference "By the Numbers: How to Analyze Race Data from Vehicle Stops," sponsored by PERF and the Office of Community Oriented Policing Services, Las Vegas, NV, July 13-14.

Colon, I. 1992. Race, Belief in Destiny, and Seat Belt Usage: A Pilot Study. *American Journal of Public Health*, 82(6): 875-877.

Cordner, Gary, Brian Williams, and Alfredo Velasco. 2002. *Vehicle Stops in San Diego: 2001*. Report submitted to the San Diego Police Department.

Cordner, Gary, Brian Williams, and Maria Zuniga. 2001. *Vehicle Stops for the Year 2000: Annual Report*. Report submitted to the San Diego Police Department.

Council on Crime and Justice and Institute on Race and Poverty. 2003. *The Minnesota Racial Profiling Report*. A report submitted to the Minnesota Legislature. Available on the PERF website at http://www.policeforum.org.

Cox, Stephen, Susan E. Pease, Daniel S. Miller, and C. Benjamin Tyson. 2001. *Interim Report of Traffic Stops Statistics for the State of Connecticut.* Report issued by the Office of the Chief State's Attorney.

Crawford, Charles E. 2000. Race and Pretextual Stops: Noise Enforcement in Midwest City. *Social Pathology*, 6(3): 213-227.

Cummings, P. 2002. Association of Seat Belt Use with Death: A Comparison of Estimates Based on Data from Police and Estimates Based on Data from Trained Crash Investigators. *Injury Prevention*, 8(4): 338-341.

Cummings, P., J. D. Wells, and F. P. Rivara. 2003. Estimating Seat Belt Effectiveness Using Matched-Pair Cohort Methods. *Accident Analysis and Prevention*, 35(1): 143-149.

Davis, Ron. 2001. *Racial Profiling: What Does the Data Mean?* Alexandria, VA: National Organization of Law Enforcement Executives.

Decker, Scott. 2002. Personal written communication, October.

Decker, Scott and Jeff Rojek. 2002. *St. Louis Metropolitan Police Department Traffic Stop Patterns*. Report submitted by the University of Missouri–St. Louis to the St. Louis Police Department, January.

Dedman, Bill and Francie Latour. 2003. Race, Sex, and Age of Drivers Ticketed. *Boston Globe*, July 20. See http://www.boston.com/globe/metro/packages/tickets/.

Deichert, Jerry. 2003. Personal written communication, June.

DeYoung, D.J., R. C. Peck, and C. J. Helander. 1997. Estimating the Exposure and Fatal Crash Rates of Suspended/Revoked and Unlicensed Drivers in California. *Crash Analysis and Prevention*, 29(1): 17-23.

Eck, John E., Lin Liu, and Lisa Growette Bostaph. 2003. *Police Vehicle Stops in Cincinnati: July 1 – December 31, 2001.* Report of parties to collaborative agreement to the monitor of that agreement.

Edwards, Terry D., Elizabeth L. Grossi, Gennaro F. Vito, and Angela D. West. 2002a. *Traffic Stop Practices of the Louisville Police Department: January 15 – December 31, 2001.* Report submitted to the Louisville Division of Police.

———. 2002b. *Traffic Stop Practices of the Iowa City Police Department: April 1 – December 31, 2001*. Report submitted to the Iowa City Police Department.

Engel, Robin Shepard, and Jennifer M. Calnon. 2003. Personal email and verbal communications, January/February.

———. 2004. Comparing Benchmark Methodologies for Police-Citizen Contacts: Traffic Stop Data Collection for the Pennsylvania State Police. *Police Quarterly*, 7(1): 97-125.

Engel, Robin Shepard, Jennifer M. Calnon, and Joshua R. Dutill. 2003. *Project on Police-Citizen Contacts, Six-Month Report*. Report prepared for the Office of the Commissioner of the Pennsylvania State Police by the Population Research Institute of the Pennsylvania State University.

Engel, Robin, Jennifer M. Calnon, Lin Liu, and Richard Johnson. 2004. *Project on Police-Citizen Contacts, Year 1 Final Report*. Report prepared for the Office of the Commissioner of the Pennsylvania State Police.

Fagan, Jeffrey. 2002. Law, Social Science, and Racial Profiling. *Justice Research and Policy*, 4(Special Issue), Fall.

Fagan, Jeffrey and Garth Davies. 2000. Street Stops and Broken Windows: Terry, Race, and Disorder in New York City. *Fordham Urban Law Journal*, 28: 457-502.

Farmer, J. 2001. Monitors' Quarterly and Training Evaluation Reports Track State Police Progress: Attorney General Releases Traffic Survey and Semiannual Data on Statewide State Police Traffic Enforcement and Trooper Conduct. Trenton, NJ: New Jersey Department of Law and Public Safety.

Farrell, Amy. 2003a. *Auditing Municipal Traffic Stop Data: Implications for Racial Profiling Analysis*. PowerPoint presentation at the conference "Confronting Racial Profiling in the Twenty-first Century: Implications for Racial Justice," sponsored by the Northeastern University Institute on Race and Justice, Boston, MA, March 8-9.

——. 2003b. Personal email communication, March.

——. 2004. *Drawing Conclusions from Results*. PowerPoint presentation at the conference "By the Numbers: How to Analyze Race Data from Vehicle Stops," sponsored by PERF and the Office of Community Oriented Policing Services, Las Vegas, NV, July 13-14.

Farrell, Amy, Jack McDevitt, Lisa Bailey, Carsten Andresen, and Erica Pierce. 2004. *Massachusetts Racial and Gender Profiling Study, Final Report.* Submitted to the Massachusetts Department of Public Safety. Available on the PERF website at http://www.policeforum.org.

Farrell, Amy, Jack McDevitt, and Michael E. Buerger. 2002. Moving Police and Community Dialogues Forward through Data Collection Task Forces. *Police Quarterly*, 5(3): 359-379.

Farrell, Amy, Jack McDevitt, Shea Cronin, and Erica Pierce. 2003. *Rhode Island Traffic Stop Statistics Act: Final Report*. Report submitted to the State of Rhode Island by the Northeastern University Institute on Race and Justice, June 30.

Feest, J. 1968. Compliance with Legal Regulations: Observation of Stop Sign Behavior. *Law and Society Review*, II: 447-461.

Friday, Steve. 2002. *Data Collection: Ohio State Highway Patrol Model*. Paper presented at the conference "Bias-Based Policing: Where Are We Now?" sponsored by the Ohio Association of Chiefs of Police, Columbus, OH, September 25.

Fridell, Lorie. 2004. *Understanding Race Data from Vehicle Stops: A Stakeholder's Guide*. Washington, D.C.: Police Executive Research Forum.

Fridell, Lorie, Robert Lunney, Drew Diamond, and Bruce Kubu. 2001. *Racially Biased Policing: A Principled Response*. Washington, D.C.: Police Executive Research Forum.

Gaines, Larry K. 2002. *An Analysis of Traffic Stop Data in the City of Riverside (California)*. Report submitted to the City of Riverside.

The Gallup Organization. 1999. Racial Profiling Is Seen as Widespread, Particularly among Young Black Men. Available at http://www.gallup.com/poll/releases/pr991209.asp.

———. 2001. Gallup Social Audit on Black/White Relations in the U.S. Available at http://www.gallup.com/poll/releases/pr010711.asp.

Glassbrenner, Donna. 2002. *Safety Belt and Helmet Use in 2002: Overall Result* (National Highway Traffic Safety Administration Technical Report [DOT HS 809 500]). Washington, D.C.: U.S. Department of Transportation, National Highway Traffic Safety Administration, September.

———. 2003. *Safety Belt Use in 2002: Demographic Characteristics* (National Highway Traffic Safety Administration Research Note [DOT HS 809 557]). Washington, D.C.: U.S. Department of Transportation, National Highway Traffic Safety Administration, March.

Greenwald, Howard P. 2001. *Final Report: Police Vehicle Stops in Sacramento California*. Report to the Sacramento Police Department, October 31. Available at http://www.sacpolice.com/report.pdf.

Gross, Samuel R. and Katherine Y. Barnes. 2002. Road Work: Racial Profiling and Drug Interdiction on the Highway. *Michigan Law Review*, 101(3): 651. Available at http://ssrn.com/abstract_id=331260.

Hagan, Frank E. 1993. *Research Methods in Criminal Justice and Criminology* (3rd ed.). Ontario: Macmillan Publishing Company.

Harcourt, Bernard E. 2003. *Rethinking Racial Profiling: A Critique of the Economics, Civil Liberties, and Constitutional Literature and of Criminal Profiling More Generally* (Public Law and Legal Theory Working Paper No. 51, 2nd Series). Chicago: University of Chicago Law School. Available at http://www.law.uchicago.edu/academics/publiclaw/index.html.

Harris, David. 1999. The Stories, the Statistics, and the Law: Why “Driving while Black” Matters. *Minnesota Law Review*, 84: 101-162.

——. 2002. *Profiles in Injustice: Why Racial Profiling Cannot Work*. New York: New Press.

Hernandez-Murillo, Ruben and John Knowles. 2004. Racial Profiling or Racist Policing? Bounds Tests in Aggregated Data. *International Economic Review*, 45(3): 959.

Institute on Race and Poverty. 2001. *Report on Traffic Stop Data Collected by the Saint Paul Police Department, April 15 through December 15, 2000.* Report submitted by the Institute on Race and Poverty to the Saint Paul Police Department.

——. 2003. *Minnesota Statewide Racial Profiling Report: All Participating Jurisdictions*. Report to the Minnesota Legislature, September 22.

Kirk, A. and N. Stamatiadis. 2001. *Evaluation of the Quasi-induced Exposure: Final Report*. Lexington, KY: University of Kentucky, Department of Civil Engineering.

Knowles, John, Nicola Persico, and Petra Todd. 2001. Racial Bias in Motor Vehicle Searches: Theory and Evidence. *Journal of Political Economy*, 109(1): 203-229.

Koornstra, M. J. 1973. A Model for Estimation of Collective Exposure and Proneness from Crash Data. *Crash Analysis and Prevention*, 5: 157-173.

Lamberth, John. 1996a. *Revised Statistical Analysis of the Incidence of Police Stops and Arrests of Black Drivers/Travelers on the New Jersey Turnpike between Interchanges 1 and 3 from the Years 1988 through 1991*. Report of defendant's expert in *State* v. *Pedro Soto*, 734 A.2d 350 (N.J. Super. Ct. Law. Div. 1996).

———. 1996b. Report of plaintiff's expert in *Wilkins* v. *Maryland State Police et al.*, Civil No. MJG-93-468 (D. Md. 1996).

———. 2001. *A Study of Biased Police Practices: Data Collection and Evaluation of the Washtenaw County Sheriff's Department.* Report submitted to the Washtenaw County (MI) Sheriff's Department, September.

———. 2002. Personal email communication, December.

———. 2003a. Personal phone and email communications, July and August.

———. 2003b. *Racial Profiling Data Analysis Study: Final Report for the San Antonio Police Department*. Report submitted to the San Antonio Police Department by Lamberth Consulting.

———. 2004. *Ann Arbor Police Department Traffic Stop Data Collection Methods and Analysis Study.* Report submitted to the Ann Arbor Police Department by Lamberth Consulting.

Lamberth, John, Karl Lamberth, and Jerry Clayton. 2003. Phone conversation, February.

Langan, Patrick A., Lawrence A. Greenfeld, Steven K. Smith, Matthew R. Durose, and David J. Levin. 2001. *Contacts between Police and the Public: Findings from the 1999 National Survey* (Report NCJ 184957). Washington, D.C.: U.S. Department of Justice.

Lange, James E., Kenneth O. Blackman, and Mark B. Johnson. 2001. *Speed Violation Survey of the New Jersey Turnpike: Final Report*. Report submitted by the Pacific Institute for Research and Evaluation to the Office of the Attorney General, State of New Jersey, December 13.

Lawson, C. T. and S. Fitzroy. 2004. Written communication, July.

Lawson, Catherine. 2003. Personal email, October 5.

Lovrich, Nicholas, Michael Gaffney, Clay Mosher, Mitchell Pickerill, and Michael R. Smith. 2003. *Washington State Patrol Stop Data Analysis Project: Data Analysis Project Report, June 1, 2003.* A report submitted to the Washington State Patrol by Washington State University. Pullman, WA: The Division of Governmental Studies and Services.

Lundman, Richard, and Robert L. Kaufman. 2003. Driving while Black: Effects of Race, Ethnicity, and Gender on Citizen Self-Reports of Traffic Stops and Police Actions. *Criminology*, 41(1): 195-220.

Lyles, R. W., P. Stamatiadis, and D. R. Lighthizer. 1991. Quasi-induced Exposure Revisited. *Crash Analysis and Prevention*, 23: 275-285.

Mastrofski, Stephen D., and Roger B. Parks. 1990. Improving Observational Studies of Police. *Criminology*, 28(August): 475-496.

Mastrofski, Stephen D., Roger B. Parks, Albert J. Reiss, Jr., Robert E. Worden, Christina DeJong, Jeffrey B. Snipes, and William Terrill. 1998. *Systematic Observation of Public Police: Applying Field Research Methods to Policy Issues* (Report NCJ 172859). Washington, D.C.: National Institute of Justice.

Mastrofski, S. D., J. B. Snipes, R. B. Parks, and C. D. Maxwell. 2000. The Helping Hand of the Law: Police Control of Citizens on Request. *Criminology*, 38: 303-342.

MVA and Joel Miller. 2000. *Profiling Populations Available for Stops and Searches* (Police Research Series Paper 131). London: Home Office.

McConnell, Elizabeth H., and Amie R. Scheidegger. 2001. *Race and Speeding Citations: Comparing Speeding Citations Issued by Air Traffic Officers with Those Issued by Ground Traffic Officers*. Paper presented at the annual meeting of the Academy of Criminal Justice Sciences, Washington, D.C., April 4-8.

McDevitt, Jack. 2003. Personal email communication, February.

McMahon, Joyce, Joel Garner, Captain Ronald Davis, and Amanda Kraus. 2002. *How to Correctly Collect and Analyze Racial Profiling Data: Your Reputation Depends on It! Final Report for Racial Profiling–Data Collection and Analysis*. Washington, D.C.: Government Printing Office. Available online at http://www.cops.usdoj.gov/Default.asp?Open=True&Item=770.

Meehan, Albert J., and Michael C. Ponder. 2002. Race and Place: The Ecology of Racial Profiling African American Motorists. *Justice Quarterly*, 19(3): 399-430.

Montgomery County (MD) Department of Police. 2001. *Traffic Stop Data Collection Analysis* (2nd report).

———. 2002. *Traffic Stop Data Collection Analysis* (3rd report).

Mosher, Clayton, Mitch Pickerill, Nicholas Lovrich, and Michael Gaffney. 2004. *The Importance of Context in Understanding Biased Policing: State Patrol Traffic Citations in Washington State*. Revision of paper presented at the 2003 meetings of the American Society of Criminology.

Moswoswe, Shamiso. 2003. Personal email communication, August.

National Center for Statistics and Analysis (NCSA). 2002. *Annual Assessment of Motor Vehicle Crashes, Safety Belts.* Washington, D.C.: U.S. Department of Transportation, National Highway Traffic Safety Administration. Available at http://www-nrd.nhtsa.dot.gov/pdf/nrd-30/NCSA/Rpts/2003/Assess02.pdf.

Norusis, Marija J. n.d. *SPSS 6.1 Guide to Data Analysis*. Englewood Cliffs, N.J.: Prentice Hall.

Novak, Kenneth. 2004. Disparity and Racial Profiling in Traffic Enforcement. *Police Quarterly*, 7(1): 65-96.

Parker, Robert. 2003. Interview regarding research in Riverside, California, June.

Parker, Robert Nash, Olivia Seheult, Kay Pih, Helen Ross, Holly Meade, Emily O'Neill, Louis Tuthill, and Bryan Wilson. n.d. *Racial Preference in Traffic Stops: A Geospatial Analysis of Driving While Black or Brown.* Available on the PERF website at http://www.policeforum.org.

Persico, Nicola. 2002. Racial Profiling, Fairness, and Effectiveness of Policing. *American Economic Review*, 92(5): 1472-1497.

Petrocelli, Matthew, Alex R. Piquero, and Michael R. Smith. 2003. Conflict Theory and Racial Profiling: An Empirical Analysis of Police Traffic Stop Data. *Journal of Criminal Justice*, 31: 1-11.

Pierchala, Carl E., and Jyoti Surti. 1999. *Control Charts as a Tool in Data Quality Improvement* (Technical report DOT HS 809 005). Washington, D.C.: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Pisarski, Alan E. 1996. *Commuting in America II: The Second National Report on Commuting Patterns and Trends*. Lansdowne, VA: Eno Transportation Foundation, Inc.

Posner, Mark A. 2002. Letter from Mark Posner, attorney with the Special Litigation Section of the Civil Rights Division of the U.S. Department of Justice, to Anthony Cowell, deputy attorney general, acting director of State Police Affairs in New Jersey, January 8.

Ramirez, D., J. McDevitt, and A. Farrell. 2000. *A Resource Guide on Racial Profiling Data Collection Systems: Promising Practices and Lessons Learned*. Washington, D.C.: U.S. Department of Justice.

Reichardt, Charles S., and Melvin M. Mark. 1998. Quasi-experimentation. In Leonard Bickman and Debra J. Rog (Eds.), *Handbook of Applied Social Research Methods*, (pp. 193-228). Thousand Oaks, CA: Sage Publications.

Reiss, Albert J., Jr. 1967. *The Police and the Public*. New Haven, CT: Yale University Press.

——. 1968. Stuff and Nonsense about Social Surveys and Observation. In Howard S. Becker, Blanche Greer, David Riesman, and Robert S. Weiss (Eds.), *Institutions and the Person: Papers Presented to C. Everett Hughes*. Chicago: Aldine Publishing Company.

——. 1971. Systematic Observation of Natural Social Phenomena. In Herbert L. Costner (Ed.), *Sociological Methodology* (pp. 3-33). San Francisco: Jossey-Bass.

Ridgeway, Greg. 2003. *Analysis of Racial Profiling in the Urban Environment*. Report for the Police Executive Research Forum by the RAND Corporation, October.

Rojek, Jeff, Richard Rosenfeld, and Scott Decker. 2002. *The Influence of Driver's Race on Traffic Stops in Missouri*. Unpublished document, University of Missouri, St. Louis.

Rudovsky, David. 2001. Law Enforcement by Stereotypes and Serendipity: Racial Profiling and Stops and Searches Without Cause. *University of Pennsylvania Journal of Constitutional Law*, 3: 298-366.

Scales, Robert. 2001. *Racial Profiling: Seattle's Community Involved Approach*. Presentation at the 2001 International Problem Oriented Policing Conference sponsored by the Police Executive Research Forum, San Diego, CA, December.

Schafer, Joseph A., David L. Carter, and Andra Katz-Bannister. 2004. Studying Traffic Stop Encounters. *Journal of Criminal Justice*, 32(2): 159-170.

Schafer, Joseph A., David L. Carter, Andra Katz-Bannister, and William M. Wells. Forthcoming. Decision-making in Traffic Stop Encounters: A Multivariate Analysis of Police Behavior. *Police Quarterly*.

Schmitt, Erica Leah, Patrick A. Langan, and Matthew R. Durose. 2002. *Characteristics of Drivers Stopped by Police, 1999* (NCJ 191548). Washington, D.C.: Bureau of Justice Statistics, U.S. Department of Justice.

Smith, Michael R. 2000. *The Traffic Stop Practices of the Richmond, Virginia, Police Department*. Final report submitted to the Richmond Police Department, November 1.

Smith, Michael R. and Geoffrey P. Alpert. 2003. Searching for Direction: Courts, Social Science, and the Adjudication of Racial Profiling Claims. *Justice Quarterly*, 19(4): 673-703.

Smith, M. and M. Petrocelli. 2001. Racial Profiling? A Multivariate Analysis of Police Traffic Stop Data. *Police Quarterly*, 4: 4-27.

Smith, Terry. 2003. *Review, Critique and Recommendations for Improving Racial Profiling Studies*. Document developed by the Service Improvement Analyst of the City of Eugene Police Department.

Smith, William R., Donald Tomaskovic-Devey, Matthew T. Zingraff, H. Marcinda Mason, Patricia Y. Warren, and Cynthia Pfaff Wright. 2003. *The North Carolina Highway Traffic Study*. Final report submitted to the National Institute of Justice, Grant No. 1999-MU-CX-0022. Washington, D.C.: National Institute of Justice.

Smith, William R., Elizabeth L. Davison, Matthew T. Zingraff, Kennon J. Rice, and Denise L. Bissler. 2004. *An Empirical Investigation of the Possible Presence and Extent of Arbitrary Profiling in the Charlotte-Mecklenburg Police Department.* Final report submitted to Charlotte-Mecklenburg Police Department.

Spitzer, Eliot. 1999. *The New York City Police Department's "Stop and Frisk" Practices: A Report to the People of the State of New York from the Office of the Attorney General*. New York: Civil Rights Bureau, Attorney General of the State of New York, December 1. Available at http://www.oag.state.ny.us/press/reports/stop_frisk/stop_frisk.html.

Stamatiadis, N., and J. A. Deacon. 1997. Quasi-induced Exposure: Methodology and Insight. *Crash Analysis and Prevention*, 29(1): 37-52.

State v. *Pedro Soto* 734 A.2d 350 (N.J. Super.Ct. Law. Div. 1996).

Thomas, Deborah. 2001. *Preliminary Summary Report: Denver Police Department Contact Card Data, June 1, 2001 through August 31, 2001*. Report provided to the Denver Police Department. Available at http://admin.denvergov.org/admin/template3/forms/DPDPreliminaryReport3monthNov2001.pdf.

——. 2002. *First Annual Report, Denver Police Department Contact Card Data Analysis, June 1, 2001 through May 31, 2002*. Report provided to the Denver Police Department, October. Available at http://admin.denvergov.org/admin/template3/forms/DPDContactCardAnnualReport102902.pdf.

Timoney, John. 2004. Panelist at "Law Enforcement Use of Force" webcast discussion sponsored by the Office of Community Oriented Policing Services at their 2004 National Community Policing Conference, June 22.

Tomaskovic-Devey, Donald, Cynthia Pfaff Wright, and Ronald Czaja. 2003. *Self-Reports of Police Speeding Stops by Race: Results from the North Carolina Reverse Record Check Survey*. Unpublished manuscript, Department of Sociology and Anthropology, North Carolina State University.

U.S. Department of Transportation, Bureau of Transportation Statistics. 2001. *Transportation Statistics Annual Report 2000* (BTS01-02). Washington, D.C.: Author. See http://www.bts.gov/publications/tsar.

U.S. Department of Transportation, Federal Highway Administration. 1995. *Nationwide Personal Transportation Survey* (Microdata Files CD-ROM). Washington, D.C.: Author.

Walker, S. and G. Alpert. 2000. *Early Warning Systems: Concept, History, and Issues* (IQ Service Report No. 8), August. Washington, D.C.: International City Management Association.

Walker, Samuel. 2001. Searching for the Denominator: Problems with Police Traffic Stop Data and an Early Warning System Solution. *Justice Research and Policy*, 3(2): 63-95.

——. 2002. *The Citizen's Guide to Interpreting Traffic Stop Data: Unraveling the Racial Profiling Controversy*. Unpublished manuscript.

——. 2003. *Internal Benchmarking for Traffic Stop Data: An Early Intervention System Approach.* Discussion paper available at http://www.policeforum.org/internalb.pdf.

Wilkins v. Maryland State Police, Settlement Agreement, Civil No. MJG-93-468 (D. Md. 1995).

Withrow, Brian L. 2002. *The Wichita Stop Study*. Report submitted to the Wichita Police Department.

Zingraff, Matthew T. 2003a. Personal email communication, March.

———.2003b. Personal phone communication, August.

Zingraff, Matthew T., H. Marcinda Mason, William R. Smith, Donald Tomaskovic-Devey, Patricia Warren, Harvey L. McMurray, and C. Robert Fenlon. 2000. *Evaluating North Carolina State Highway Patrol Data: Citations, Warnings and Searches in 1998*. Report submitted to the North Carolina Department of Crime Control and Public Safety and North Carolina State Highway Patrol. Available online at http://www.nccrimecontrol.org/shp/ncshpreport.htm.

Zingraff, Matthew T., William R. Smith, and Donald Tomaskovic-Devey. 2001. *What Traffic Stop Data Can Tell Us About Possible Racial Bias in Policing: Results from the North Carolina State Highway Patrol Study*. Presentation at the Annual Meeting of the American Society of Criminology. Atlanta, GA.

About the Author

Dr. Lorie Fridell is Director of Research for the Police Executive Research Forum (PERF) and a social scientist by training. Prior to joining PERF in 1999, she was a tenured associate professor of criminology and criminal justice first at the University of Nebraska and then at Florida State University. She has been conducting research on law enforcement for more than 15 years and is a national expert on racial profiling. The lead author of *Racially Biased Policing: A Principled Response* (PERF 2001), Fridell also has written extensively on such topics as police use of force, citizen complaints, police pursuits, and problem-oriented policing.

About the Office of Community Oriented Policing Services (COPS) U.S. Department of Justice

The Office of Community Oriented Policing Services (COPS) was created in 1994 and has the unique mission to directly serve the needs of state and local law enforcement. The COPS Office has been the driving force in advancing the concept of community policing, and is responsible for one of the greatest infusions of resources into state and local law enforcement in our nation's history.

Since 1994, COPS has invested $10.6 billion to add community policing officers to the nation's streets, enhance crime fighting technology, support crime prevention initiatives, and provide training and technical assistance to help advance community policing. COPS funding has furthered the advancement of community policing through community policing innovation conferences, the development of best practices, pilot community policing programs, and applied research and evaluation initiatives. COPS has also positioned itself to respond directly to emerging law enforcement needs. Examples include working in partnership with departments to enhance police integrity, promoting safe schools, and combating the methamphetamine drug problem.

The COPS Office has made substantial investments in law enforcement training. COPS created a national network of Regional Community Policing Institutes that are available to state and local law enforcement, elected officials and community leaders for training opportunities on a wide range of community policing topics. COPS also supports the advancement of community policing strategies through the Community Policing Consortium. Additionally, COPS has made a major investment in applied research which makes possible the growing body of substantive knowledge covering all aspects of community policing.

These substantial investments have produced a significant national community policing infrastructure, as evidenced by the fact that at the present time, approximately 86% of the nation's population is served by law enforcement agencies practicing community policing. The COPS Office continues to respond proactively by providing critical resources, training, and technical assistance to help state and local law enforcement implement innovative and effective community policing strategies.

About PERF

The Police Executive Research Forum (PERF) is a national professional association of chief executives of large city, county and state law enforcement agencies. PERF's objective is to improve the delivery of police services and the effectiveness of crime control through several means:

- the exercise of strong national leadership,
- the public debate of police and criminal justice issues,
- the development of research and policy, and
- the provision of vital management and leadership services to police agencies.

PERF members are selected on the basis of their commitment to PERF's objectives and principles. PERF operates under the following tenets:

- Research, experimentation and exchange of ideas through public discussion and debate are paths for the development of a comprehensive body of knowledge about policing.
- Substantial and purposeful academic study is a prerequisite for acquiring, understanding and adding to that body of knowledge.
- Maintenance of the highest standards of ethics and integrity is imperative in the improvement of policing.
- The police must, within the limits of the law, be responsible and accountable to citizens as the ultimate source of police authority.
- The principles embodied in the Constitution are the foundation of policing.